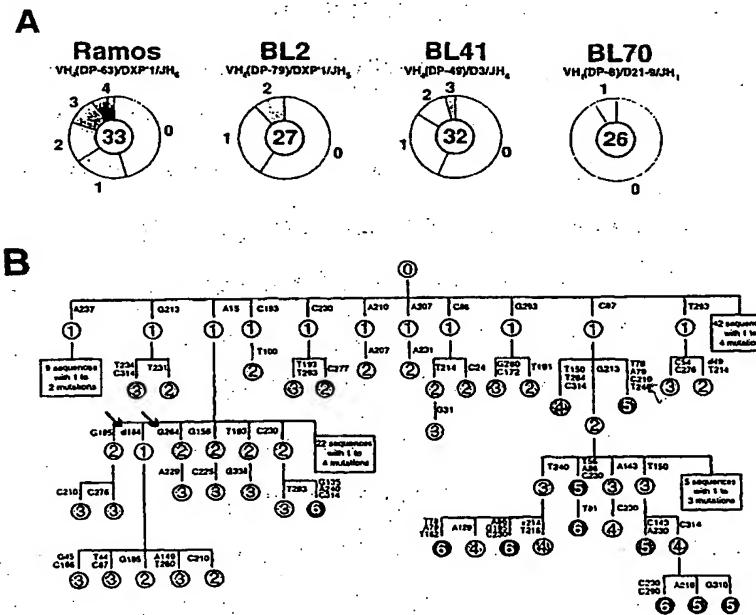




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/10, 5/08, 5/10, C07K 16/00, C12Q 1/68, A01K 67/027		A1	(11) International Publication Number: WO 00/22111
			(43) International Publication Date: 20 April 2000 (20.04.00)
(21) International Application Number: PCT/GB99/03358		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 8 October 1999 (08.10.99)			
(30) Priority Data: 9822104.7 9 October 1998 (09.10.98) GB 9901141.3 19 January 1999 (19.01.99) GB 9913435.5 9 June 1999 (09.06.99) GB			
(71) Applicant (for all designated States except US): MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): SALE, Julian, Edward [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). NEUBERGER, Michael, Samuel [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). CUMBERS, Sarah, Jane [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB).			
(74) Agents: MASCHIO, Antonio et al.; D Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).			

(54) Title: METHOD FOR GENERATING DIVERSITY.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	México	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

Method for Generating Diversity

The present invention relates to a method for generating diversity in a gene or gene product by exploiting the natural somatic hypermutation capability of antibody-producing cells, as well as to cell lines capable of generating diversity in defined gene products.

Many *in vitro* approaches to the generation of diversity in gene products rely on the generation of a very large number of mutants which are then selected using powerful selection technologies. For example, phage display technology has been highly successful as providing a vehicle that allows for the selection of a displayed protein (Smith, 1985; Bass *et al.*, 1990; McCafferty *et al.*, 1990; for review see Clackson and Wells, 1994). Similarly, specific peptide ligands have been selected for binding to receptors by affinity selection using large libraries of peptides linked to the C terminus of the lac repressor LacI (Cull *et al.*, 1992). When expressed in *E. coli* the repressor protein physically links the ligand to the encoding plasmid by binding to a lac operator sequence on the plasmid. Moreover, an entirely *in vitro* polysome display system has also been reported (Mattheakis *et al.*, 1994) in which nascent peptides are physically attached via the ribosome to the RNA which encodes them.

In vivo the primary repertoire of antibody specificities is created by a process of DNA rearrangement involving the joining of immunoglobulin V, D and J gene segments. Following antigen encounter in mouse and man, the rearranged V genes in those B cells that have been triggered by the antigen are subjected to a second wave of diversification, this time by somatic hypermutation. This hypermutation generates the secondary repertoire from which good binding specificities can be selected thereby allowing affinity maturation of the humoral immune response.

Artificial selection systems to date rely heavily on initial mutation and selection, similar in concept to the initial phase of V-D-J rearrangement which occurs in natural antibody

production, in that it results in the generation of a "fixed" repertoire of gene product mutants from which gene products having the desired activity may be selected.

5 *In vitro* RNA selection and evolution (Ellington and Szostak, 1990), sometimes referred to as SELEX (systematic evolution of ligands by exponential enrichment) (Tuerk and Gold, 1990) allows for selection for both binding and chemical activity, but only for nucleic acids. When selection is for binding, a pool of nucleic acids is incubated with immobilised substrate. Non-binders are washed away, then the binders are released, amplified and the whole process is repeated in iterative steps to enrich for better binding 10 sequences. This method can also be adapted to allow isolation of catalytic RNA and DNA (Green and Szostak, 1992; for reviews see Chapman and Szostak, 1994; Joyce, 1994; Gold *et al.*, 1995; Moore, 1995). SELEX, thus, permits cyclical steps of improvement of the desired activity, but is limited in its scope to the preparation of nucleic acids.

15 Unlike in the natural immune system, however, artificial selection systems are poorly suited to any facile form of "affinity maturation", or cyclical steps of repertoire generation and development. One of the reasons for this is that it is difficult to target mutations to regions of the molecule where they are required, so subsequent cycles of 20 mutation and selection do not lead to the isolation of molecules with improved activity with sufficient efficiency.

25 Much of what is known about the somatic hypermutation process which occurs during affinity maturation in natural antibody production has been derived from an analysis of the mutations that have occurred during hypermutation *in vivo* (for reviews see Neuberger and Milstein, 1995; Weill and Raynaud, 1996; Parham, 1998). Most of these mutations are single nucleotide substitutions which are introduced in a stepwise manner. They are scattered over the rearranged V domain, though with characteristic hotspots, and the substitutions exhibit a bias for base transitions. The mutations largely 30 accumulate during B cell expansion in germinal centres (rather than during other stages of B cell differentiation and proliferation) with the rate of incorporation of nucleotide

substitutions into the V gene during the hypermutation phase estimated at between 10^{-4} and 10^{-3} bp $^{-1}$ generation $^{-1}$ (McKean et al., 1984; Berek & Milstein, 1988)

The possibility that lymphoid cell lines could provide a tractable system for investigating hypermutation was considered many years ago (Coffino and Scharff, 1971; Adetugbo et al., 1977; Brüggemann et al., 1982). Clearly, it is important that the rate of V gene mutation in the cell-line under study is sufficiently high not only to provide a workable assay but also to be confident that mutations are truly generated by the localised antibody hypermutation mechanism rather than reflecting a generally increased mutation rate as is characteristically associated with many tumours. Extensive studies on mutation have been performed monitoring the reversion of stop codons in V_H in mouse pre-B and plasmacytoma cell lines (Wabl et al., 1985; Chui et al., 1995; Zhu et al., 1995; reviewed by Green et al., 1998). The alternative strategy of direct sequencing of the expressed V gene has indicated that V_H gene diversification in several follicular, Burkitt and Hodgkin lymphomas can continue following the initial transformation event (Bahler and Levy, 1992; Jain et al., 1994; Chapman et al., 1995 and 1996; Braeuninger et al., 1997). Direct sequencing has also revealed a low prevalence of mutations in a cloned follicular lymphoma line arguing that V_H diversification can continue in vitro (Wu et al., 1995). None of the reports of constitutive mutation in cell lines cited above provides evidence that the mutations seen are the result of directed hypermutation, as observed in natural antibody diversification, which is concentrated in the V genes, as opposed to a general susceptibility to mutation as described in many tumour cell lines from different lineages.

Recently, hypermutation has been induced in a cell line by Denepoux et al. (1997), by culturing cells in the presence of anti-immunoglobulin antibody and activated T-cells. However, the hypermutation observed was stated to be induced, not constitutive.

Summary of the Invention

In a first aspect of the invention there is provided a method for preparing a lymphoid cell line capable of directed constitutive hypermutation of a target nucleic acid region, 5 comprising screening a cell population for ongoing target sequence diversification, and selecting a cell in which the rate of target nucleic acid mutation exceeds that of other nucleic acid mutation by a factor of 100 or more.

As used herein, "directed constitutive hypermutation" refers to the ability, observed for 10 the first time in experiments reported herein, of certain cell lines to cause alteration of the nucleic acid sequence of one or more specific sections of endogenous or transgene DNA in a constitutive manner, that is without the requirement for external stimulation. In cells capable of directed constitutive hypermutation, sequences outside of the specific 15 sections of endogenous or transgene DNA are not subjected to mutation rates above background mutation rates.

A "target nucleic acid region" is a nucleic acid sequence or region in the cell according to the invention which is subjected to directed constitutive hypermutation. The target nucleic acid may comprise one or more transcription units encoding gene products, 20 which may be homologous or heterologous to the cell. Exemplary target nucleic acid regions are immunoglobulin V_h genes as found in immunoglobulin-producing cells. These genes are under the influence of hypermutation-recruiting elements, as described further below, which direct the hypermutation to the locus in question. Other target 25 nucleic acid sequences may be constructed, for example by replacing V_h gene transcription units in loci which contain hypermutation-recruiting elements with another desired transcription unit, or by constructing artificial genes comprising hypermutation-recruiting elements.

"Hypermutation" refers to the mutation of a nucleic acid in a cell at a rate above 30 background. Preferably, hypermutation refers to a rate of mutation of between 10^{-5} and 10^{-3} bp⁻¹ generation⁻¹. This is greatly in excess of background mutation rates, which are

of the order of 10^{-9} to 10^{-10} mutations bp^{-1} generation $^{-1}$ (Drake *et al.*, 1988) and of spontaneous mutations observed in PCR. 30 cycles of amplification with Pfu polymerase would produce $<0.05 \times 10^{-3}$ mutations bp^{-1} in the product, which in the present case would account for less than 1 in 100 of the observed mutations (Lundberg *et al.*, 1991).

Hypermutation is a part of the natural generation of immunoglobulin variable chain (V) genes. According to the present invention therefore, the cell line is preferably an immunoglobulin-producing cell line which is capable of producing at least one immunoglobulin V gene. A V gene may be a variable light chain (V_L) or variable heavy chain (V_H) gene, and may be produced as part of an entire immunoglobulin molecule; it may be a V gene from an antibody, a T-cell receptor or another member of the immunoglobulin superfamily. Members of the immunoglobulin superfamily are involved in many aspects of cellular and non-cellular interactions *in vivo*, including widespread roles in the immune system (for example, antibodies, T-cell receptor molecules and the like), involvement in cell adhesion (for example the ICAM molecules) and intracellular signalling (for example, receptor molecules, such as the PDGF receptor). Thus, preferred cell lines according to the invention are derived from B-cells. According to the present invention, it has been determined that cell lines derived from antibody-producing B cells may be isolated which retain the ability to hypermutate V region genes, yet do not hypermutate other genes.

In a preferred embodiment, the cells according to the invention are derived from or related to cells which hypermutate *in vivo*. Cells which hypermutate *in vivo* are, for example, immunoglobulin-expressing cells, such as B-cells. Lymphoma cells, which are Ig-expressing cell tumours, are particularly good candidates for the isolation of constitutively hypermutating cell lines according to the present invention.

As used herein, "screening for ongoing target sequence diversification" refers to the determination of the presence of hypermutation in the target nucleic acid region of the cell lines being tested. This can be performed in a variety of ways, including direct

sequencing or indirect methods such as the MutS assay (Jolly *et al.*, 1997) or monitoring the generation of immunoglobulin loss variants. Cells selected according to this procedure are cells which display target sequence diversification.

5 The cell population which is subjected to selection by the method of the invention may be a polyclonal population, comprising a variety of cell types and/or a variety of target sequences, or a (mono-)clonal population of cells.

10 A clonal cell population is a population of cells derived from a single clone, such that the cells would be identical save for mutations occurring therein. Use of a clonal cell population preferably excludes co-culturing with other cell types, such as activated T-cells, with the aim of inducing V gene hypermutation.

15 Cells according to the invention do not rely on the use of induction steps in order to produce hypermutation.

Preferably, the clonal cell population screened in the present invention is derived from a B cell. Advantageously it is a lymphoma cell line, such as a Burkitt lymphoma cell line, a follicular lymphoma cell line or a diffuse large cell lymphoma cell line.

20 Preferably, the method according to the invention further comprises the steps of isolating one or more cells which display target sequence diversification, and comparing the rate of accumulation of mutations in the target sequences with that in non-target sequences in the isolated cells.

25 A feature of the present invention is that the hypermutation is directed only to specific (target) nucleic acid regions, and is not observed outside of these regions in a general manner. Specificity is thus assayed as part of the method of the invention by assaying the rate of mutation of sequences other than target sequences. C region genes, which are not naturally exposed to hypermutation, may advantageously be employed in such a technique, although any other nucleic acid region not subject to specific hypermutation

may also be used. Since hypermutation is not sequence dependent, the actual sequence of the nucleic acid region selected for comparison purposes is not important. However, it must not be subject to control sequences which direct hypermutation, as described below. Conveniently, background mutation may be assessed by fluctuation analysis, for example at the HPRT locus [see Luria and Delbreck, (1943); Capizzi and Jameson, (1973)].

Cells in which target region mutation exceeds non-target region mutation are cells capable of directed constitutive hypermutation of a specific nucleic acid region in accordance with the present invention. The factor by which V region gene mutation exceeds other gene mutation is variable, but is, in general of the order of at least 10^2 , advantageously 10^3 , and preferably 10^4 or more.

Overall mutation rates and diversity may be increased, for example by the administration of mutagens or expression of sequence modifying genes, such as terminal deoxynucleotidyl transferase (TdT). However, the difference between hypermutation and background is not expected to be increased in such a manner.

In a second aspect of the present invention, there is provided a method for preparing a gene product having a desired activity, comprising the steps of:

- a) expressing a nucleic acid encoding the gene product in a population of cells according to the first aspect of the present invention, operably linked to a nucleic acid which directs hypermutation;
- b) identifying a cell or cells within the population of cells which expresses a mutant gene product having the desired activity; and
- c) establishing one or more clonal populations of cells from the cell or cells identified in step (b), and selecting from said clonal populations a cell or cells which expresses a gene product having an improved desired activity.

30 The population of cells according to part a) above is derived from a clonal or polyclonal population of cells which comprises cells identified by a method according to the first

aspect of the invention as being capable of constitutive hypermutation of V region genes. The gene product may thus be the endogenous immunoglobulin polypeptide, a gene product expressed by a manipulated endogenous gene or a gene product expressed by a heterologous transcription unit operatively linked to control sequences which direct 5 somatic hypermutation, as described further below.

The nucleic acid which is expressed in the cells of the invention and subjected to hypermutation may be an endogenous region, such as the endogenous V region, or a heterologous region inserted into the cell line of the invention. This may take form, for 10 example, of a replacement of the endogenous V region with heterologous transcription unit(s), such as a heterologous V region, retaining the endogenous control sequences which direct hypermutation; or of the insertion into the cell of a heterologous transcription unit under the control of its own control sequences to direct hypermutation, wherein the transcription unit may encode V region genes or any other 15 desired gene product. The nucleic acid according to the invention is described in more detail below.

In step b) above, the cells are screened for the desired gene product activity. This may be, for example in the case of immunoglobulins, a binding activity. Other activities 20 may also be assessed, such as enzymatic activities or the like, using appropriate assay procedures. Where the gene product is displayed on the surface of the cell, cells which produce the desired activity may be isolated by detection of the activity on the cell surface, for example by fluorescence, or by immobilising the cell to a substrate via the 25 surface gene product. Where the activity is secreted into the growth medium, or otherwise assessable only for the entire cell culture as opposed to in each individual cell, it is advantageous to establish a plurality of clonal populations from step a) in order to increase the probability of identifying a cell which secretes a gene product having the desired activity. Advantageously, the selection system employed does not affect the cell's ability to proliferate and mutate.

Preferably, at this stage (and in step c) cells which express gene products having a better, improved or more desirable activity are selected. Such an activity is, for example, a higher affinity binding for a given ligand, or a more effective enzymatic activity. Thus, the method allows for selection of cells on the basis of a qualitative and/or quantitative assessment of the desired activity.

In a third aspect of the present invention, there is provided the use of a cell capable of directed constitutive hypermutation of a specific nucleic acid region in the preparation of a gene product having a desired activity.

In the use according to the invention, a nucleic acid encoding the gene product having the desired activity is operatively linked to control sequences which direct hypermutation within the cell. Successive generations of the cell thus produce mutants of the nucleic acid sequence, which are screened by the method of the invention to isolate mutants with advantageous properties.

Brief Description of the Figures

Figure 1. V_H diversity in Burkitt lines.

(A) Sequence diversity in the rearranged V_H genes of four sporadic Burkitt lymphoma lines, shown as pie charts. The number of M13 clones sequenced for each cell line is denoted in the centre of the pie; the sizes of the various segments depict the proportion of sequences that are distinguished by 0, 1, 2 etc. mutations (as indicated) from the consensus.

(B) Presumed dynastic relationship of V_H mutations identified in the initial Ramos culture. Each circle (with shading proportional to extent of mutation) represents a distinct sequence with the number of mutations accumulated indicated within the circle.

(C) Mutation prevalence in the rearranged V_λ genes. Two V_λ rearrangements are identified in Ramos. Diversity and assignment of germline origin is presented as in Figure 1A.

(D) Comparison of mutation prevalence in the V_H and $C\mu$ regions of the initial Ramos culture. Pie charts are presented as in Figure 1A.

5 Figure 2 Constitutive V_H diversification in Ramos.

(A) Diversification assessed by a MutS assay. The mutation prevalence in each population as deduced by direct cloning and sequencing is indicated.

(B) Dynastic relationships deduced from the progeny of three independent Ramos clones.

10

Figure 3. Distribution of unselected nucleotide substitutions along the Ramos V_H .

Figure 4. Hypermutation in Ramos generates diverse revertible IgM-loss variants.

(A) Scheme showing the isolation of IgM-loss variants.

(B) Table showing that multiple nonsense mutations can contribute to V_H inactivation. Each V_H codon position at which stops are observed in these two populations is listed.

(C) Table of reversion rates of IgM-loss variants.

(D) Sequence surrounding the stop codons in the IgM-loss derivatives.

20

Figure 5. IgM-loss variants in Ramos transfectants expressing TdT.

(A) Western blot analysis of expression of TdT in three pSV-p β G/TdT and three control transfectants of Ramos.

(B) Pie charts depicting independent mutational events giving rise to IgM-loss variants.

25

Figure 6. Sequence table summarising mutations in V_H other than single nucleotide substitutions.

30 Figure 7. Comparison of sequences isolated from V_H genes of Ramos cells which have lost anti-idiotype (anti-Id1) binding specificity. Nucleotide substitutions which

differ from the starting population consensus are shown in bold. Predicted amino acid changes are indicated, also in bold type.

Figure 8. Bar graph showing enrichment of Ramos cells for production of an immunoglobulin with a novel binding specificity, by iterative selection over five rounds.

Figure 9. Bar graph showing improved recovery of Ramos cells binding a novel specificity (streptavidin) by increasing the bead:cell ratio.

Figure 10. Chart showing increase in recovery of novel binding specificity Ramos cells according to increasing target antigen concentration.

Figure 11. V_H sequence derived from streptavidin-binding Ramos cells. Nucleotide changes observed in comparison with the V_H sequence of the starting population, and predicted amino acid changes, are shown in bold.

Figure 12. Amount of IgM in supernatants of cells selected in rounds 4, 6 and 7 of a selection process for streptavidin binding, against control medium and unselected Ramos cell supernatant.

Figure 13. Streptavidin binding of IgM from the supernatants of Figure 12.

Figure 14. Streptavidin binding of supernatants from round 4 and round 6 of a selection for streptavidin binding, analysed by surface plasmon resonance.

Figure 15. FACS analysis of binding to streptavidin-FITC of cells selected in rounds 4 and 6.

Figure 16. V_H and V_L sequences of round 6 selected IgM.

Detailed Description of the Invention

The present invention makes available for the first time a cell line which constitutively hypermutates selected nucleic acid regions. This permits the design of systems which

5 produce mutated gene products by a technique which mirrors affinity maturation in natural antibody production. The Ramos Burkitt line constitutively diversifies its rearranged immunoglobulin V gene during in vitro culture. This hypermutation does not require stimulation by activated T cells, exogenously-added cytokines or even maintenance of the B cell antigen receptor.

10

The rate of mutation (which lies in the range $0.2-1 \times 10^{-4}$ bp $^{-1}$ generation $^{-1}$) is sufficiently high to readily allow the accumulation of a large database of unselected mutations and so reveal that hypermutation in Ramos exhibits most of the features classically associated with immunoglobulin V gene hypermutation in vivo (preferential targeting of mutation to the V; stepwise accumulation of single nucleotide substitutions; transition bias; characteristic mutational hotspots). The large majority of mutations in the unselected database are single nucleotide substitutions although deletions and duplications (sometimes with a flanking nucleotide substitution) are detectable. Such deletions and duplications have also been proposed to be generated as a consequence of hypermutation in vivo (Wilson et al., 1998; Goosens et al., 1998; Wu & Kaartinen, 20 1995).

25 The isolation of cells which constitutively hypermutate selected nucleic acid regions is based on the monitoring of V gene mutation in cell lines derived from antibody-producing cells such as B cells. The selection method employed in the invention may be configured in a number of ways.

Selection of Hypermutating Cells

30 Hypermutating cells may be selected from a population of cells by a variety of techniques, including sequencing of target sequences, selection for expression loss

mutants, assay using bacterial MutS protein and selection for change in gene product activity.

One of the features of hypermutation of target nucleic acids is that the process results in the introduction of stop codons into the target sequence with far greater frequency than would be observed in the absence of hypermutation. This results in loss of production of a gene product from the cell. This loss may be exploited to identify cells which are hypermutating nucleic acid sequences.

In a preferred embodiment of the invention, the target nucleic acid encodes an immunoglobulin. Immunoglobulin loss may be detected both for cells which secrete immunoglobulins into the culture medium, and for cells in which the immunoglobulin is displayed on the cell surface. Where the immunoglobulin is present on the cell surface, its absence may be identified for individual cells, for example by FACS analysis, immunofluorescence microscopy or ligand immobilisation to a support. In a preferred embodiment, cells may be mixed with antigen-coated magnetic beads which, when sedimented, will remove from the cell suspension all cells having an immunoglobulin of the desired specificity displayed on the surface.

The technique may be extended to any immunoglobulin molecule, including antibodies, T-cell receptors and the like. The selection of immunoglobulin molecules will depend on the nature of the clonal population of cells which it is desired to assay according to the invention.

Alternatively, cells according to the invention may be selected by sequencing of target nucleic acids, such as V genes, and detection of mutations by sequence comparison. This process may be automated in order to increase throughput.

In a further embodiment, cells which hypermutate V genes may be detected by assessing change in antigen binding activity in the immunoglobulins produced in a clonal cell population. For example, the quantity of antigen bound by a specific unit amount of

cell medium or extract may be assessed in order to determine the proportion of immunoglobulin produced by the cell which retains a specified binding activity. As the V genes are mutated, so binding activity will be varied and the proportion of produced immunoglobulin which binds a specified antigen will be reduced.

5

Alternatively, cells may be assessed in a similar manner for the ability to develop a novel binding affinity, such as by exposing them to an antigen or mixture of antigens which are initially not bound and observing whether a binding affinity develops as the result of hypermutation.

10

In a further embodiment, the bacterial MutS assay may be used to detect sequence variation in target nucleic acids. The MutS protein binds to mismatches in nucleic acid hybrids. By creating heteroduplexes between parental nucleic acids and those of potentially mutated progeny, the extent of mismatch formation, and thus the extent of nucleic acid mutation, can be assessed.

15

Where the target nucleic acid encodes a gene product other than an immunoglobulin, selection may be performed by screening for loss or alteration of a function other than binding. For example, the loss or alteration of an enzymatic activity may be screened for.

20

Cells which target sequence hypermutation are assessed for mutation in other nucleic acid regions. A convenient region to assay is the constant (C) region of an immunoglobulin gene. C regions are not subject to directed hypermutation according to the invention. The assessment of C regions is preferably made by sequencing and comparison, since this is the most certain method for determining the *absence* of mutations. However, other techniques may be employed, such as monitoring for the retention of C region activities, for example complement fixation, which may be disrupted by hypermutation events.

25

30

Adaptation of the endogenous gene products

Having obtained a cell line which constitutively hypermutates an endogenous gene, such as an immunoglobulin V-region gene, the present invention provides for the adaptation 5 of the endogenous gene product, by constitutive hypermutation, to produce a gene product having novel properties. For example, the present invention provides for the production of an immunoglobulin having a novel binding specificity or an altered binding affinity.

10 The process of hypermutation is employed, in nature, to generate improved or novel binding specificities in immunoglobulin molecules. Thus, by selecting cells according 15 to the invention which produce immunoglobulins capable of binding to the desired antigen and then propagating these cells in order to allow the generation of further mutants, cells which express immunoglobulins having improved binding to the desired 20 antigen may be isolated.

A variety of selection procedures may be applied for the isolation of mutants having a desired specificity. These include Fluorescence Activated Cell Sorting (FACS), cell separation using magnetic particles, antigen chromatography methods and other cell 25 separation techniques such as use of polystyrene beads.

Separating cells using magnetic capture may be accomplished by conjugating the antigen of interest to magnetic particles or beads. For example, the antigen may be conjugated 20 to superparamagnetic iron-dextran particles or beads as supplied by Miltenyi Biotec GmbH. These conjugated particles or beads are then mixed with a cell population which may express a diversity of surface immunoglobulins. If a particular cell 25 expresses an immunoglobulin capable of binding the antigen, it will become complexed with the magnetic beads by virtue of this interaction. A magnetic field is then applied 30 to the suspension which immobilises the magnetic particles, and retains any cells which are associated with them via the covalently linked antigen. Unbound cells which do not become linked to the beads are then washed away, leaving a population of cells which is

isolated purely on its ability to bind the antigen of interest. Reagents and kits are available from various sources for performing such one-step isolations, and include Dynal Beads (Dynal AS; <http://www.dynal.no>), MACS-Magnetic Cell Sorting (Miltenyi Biotec GmbH; <http://www.miltenyibiotec.com>), ClinIMACS (AmCell; <http://www.amcell.com>) as well as Biomag, Amerlex-M beads and others.

Fluorescence Activated Cell Sorting (FACS) can be used to isolate cells on the basis of their differing surface molecules, for example surface displayed immunoglobulins. Cells in the sample or population to be sorted are stained with specific fluorescent reagents which bind to the cell surface molecules. These reagents would be the antigen(s) of interest linked (either directly or indirectly) to fluorescent markers such as fluorescein, Texas Red, malachite green, green fluorescent protein (GFP), or any other fluorophore known to those skilled in the art. The cell population is then introduced into the vibrating flow chamber of the FACS machine. The cell stream passing out of the chamber is encased in a sheath of buffer fluid such as PBS (Phosphate Buffered Saline). The stream is illuminated by laser light and each cell is measured for fluorescence, indicating binding of the fluorescent labelled antigen. The vibration in the cell stream causes it to break up into droplets, which carry a small electrical charge. These droplets can be steered by electric deflection plates under computer control to collect different cell populations according to their affinity for the fluorescent labelled antigen. In this manner, cell populations which exhibit different affinities for the antigen(s) of interest can be easily separated from those cells which do not bind the antigen. FACS machines and reagents for use in FACS are widely available from sources world-wide such as Becton-Dickinson, or from service providers such as Arizona Research Laboratories (<http://www.arl.arizona.edu/facs/>).

Another method which can be used to separate populations of cells according to the affinity of their cell surface protein(s) for a particular antigen is affinity chromatography. In this method, a suitable resin (for example CL-600 Sepharose, Pharmacia Inc.) is covalently linked to the appropriate antigen. This resin is packed into a column, and the mixed population of cells is passed over the column. After a

suitable period of incubation (for example 20 minutes), unbound cells are washed away using (for example) PBS buffer. This leaves only that subset of cells expressing immunoglobulins which bound the antigen(s) of interest, and these cells are then eluted from the column using (for example) an excess of the antigen of interest, or by 5 enzymatically or chemically cleaving the antigen from the resin. This may be done using a specific protease such as factor X, thrombin, or other specific protease known to those skilled in the art to cleave the antigen from the column via an appropriate cleavage site which has previously been incorporated into the antigen-resin complex. Alternatively, a non-specific protease, for example trypsin, may be employed to remove 10 the antigen from the resin, thereby releasing that population of cells which exhibited affinity for the antigen of interest.

Insertion of heterologous transcription units

15 In order to maximise the chances of quickly selecting an antibody variant capable of binding to any given antigen, or to exploit the hypermutation system for non-immunoglobulin genes, a number of techniques may be employed to engineer cells according to the invention such that their hypermutating abilities may be exploited.

20 In a first embodiment, transgenes are transfected into a cell according to the invention such that the transgenes become targets for the directed hypermutation events.

As used herein, a "transgene" is a nucleic acid molecule which is inserted into a cell, 25 such as by transfection or transduction. For example, a "transgene" may comprise a heterologous transcription unit as referred to above, which may be inserted into the genome of a cell at a desired location.

The plasmids used for delivering the transgene to the cells are of conventional 30 construction and comprise a coding sequence, encoding the desired gene product, under the control of a promoter. Gene transcription from vectors in cells according to the

invention may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with the heterologous coding sequence, provided such promoters are compatible with the host system of the invention.

Transcription of a heterologous coding sequence by cells according to the invention may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to the coding sequence, but is preferably located at a site 5' from the promoter.

Advantageously, a eukaryotic expression vector may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the heterologous coding sequence is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, in vectors designed for gene therapy applications or in transgenic animals.

Eukaryotic expression vectors will also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA.

An expression vector includes any vector capable of expressing a coding sequence encoding a desired gene product that is operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a 5 phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well-known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding a heterologous 10 coding sequence may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., 1989).

Construction of vectors according to the invention employs conventional ligation 15 techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing gene product expression and 20 function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the 25 art will readily envisage how these methods may be modified, if desired.

In one variation of the first embodiment, transgenes according to the invention also 30 comprise sequences which direct hypermutation. Such sequences have been characterised, and include those sequences set forth in Klix *et al.*, (1998), and Sharpe *et*

encoding the gene product is transferred into the cells. The transcription unit and the sequences which direct hypermutation are thus exogenous to the cell. However, although exogenous the sequences which direct hypermutation themselves may be similar or identical to the sequences which direct hypermutation naturally found in the cell.

In a second embodiment, the endogenous V gene(s) or segments thereof may be replaced with heterologous V gene(s) by homologous recombination, or by gene targeting using, for example, a Lox/Cre system or an analogous technology or by insertion into hypermutating cell lines which have spontaneously deleted endogenous V genes. Alternatively, V region gene(s) may be replaced by exploiting the observation that hypermutation is accompanied by double stranded breaks in the vicinity of rearranged V genes.

The invention is further described below, for the purposes of illustration only, in the following examples.

Example 1: Selection of a hypermutating cell

In order to screen for a cell that undergoes hypermutation in vitro, the extent of diversity that accumulates in several human Burkitt lymphomas during clonal expansion is assessed. The Burkitt lines BL2, BL41 and BL70 are kindly provided by G. Lenoir (IARC, Lyon, France) and Ramos (Klein et al., 1975) is provided by D. Fearon (Cambridge, UK). Their rearranged V_H genes are PCR amplified from genomic DNA using multiple V_H family primers together with a J_H consensus oligonucleotide. Amplification of rearranged V_H segments is accomplished using Pfu polymerase together with one of 14 primers designed for each of the major human V_H families (Tomlinson, 1997) and a consensus J_H back primer which anneals to all six human J_H segments (JOL48, 5'-GCGGTACCTGAGGAGACGGTGACC-3', gift of C. Jolly). Amplification of the Ramos V_H from genomic DNA is performed with oligonucleotides RVHFOR (5'-CCCCAAGCTTCCCAGGTGCAGCTACAGCAG) and JOL48. Amplification of

the expressed V_H - $C\mu$ cDNA is performed using RVHFOR and $C\mu$ 2BACK (5'-CCCGGTACCAAGATGAGCTTGGACTTGCAG). The genomic $C\mu$ 1/2 region is amplified using $C\mu$ 2BACK with $C\mu$ 1FOR (5'-CCCAAGCTTCAGTCTGCCCTGACTCAG) and 5'-CCCTCTAGACCACCTAGGACGGTCAGCTT. PCR products are purified using QIAquick (Qiagen) spin columns and sequenced using an ABI377 sequencer following cloning into M13. Mutations are computed using the GAP4 alignment program (Bonfield et al., 1995).

Sequencing of the cloned PCR products reveals considerable diversity in the Ramos cell line (a prevalence of 2.8×10^{-3} mutations bp⁻¹ in the V_H) although significant heterogeneity is also observed in BL41 as well as in BL2. See Figure 1A. Sequence diversity in the rearranged V_H genes of four sporadic Burkitt lymphoma lines are shown as pie charts. The rearranged V_H genes in each cell line are PCR amplified and cloned into M13. For each cell line, the consensus is taken as the sequence common to the greatest number of M13 clones and a germline counterpart (indicated above each pie) assigned on the basis of closest match using the VBASE database of human immunoglobulin sequences (Tomlinson, 1997). The V_H consensus sequence for Ramos used herein differs in 3 positions from the sequence determined by Chapman et al (1996), five positions from that determined by Ratech (1992) and six positions from its closest germline counterpart V_H 4(DP-63).

The analysis of V_H diversity in Ramos is extended by sequencing the products from nine independent PCR amplifications. This enables a likely dynastic relationship between the mutated clones in the population to be deduced, minimising the number of presumed independent repeats of individual nucleotide substitutions (Figure 1B). 315 M13 V_H clones obtained from nine independent PCR amplifications are sequenced; the dynasty only includes sequences identified (rather than presumed intermediates). Individual mutations are designated according to the format "C230" with 230 being the nucleotide position in the Ramos V_H (numbered as in Figure 3) and the "C" indicating

the novel base at that position. The criterion used to deduce the genealogy is a minimisation of the number of independent occurrences of the same nucleotide substitution. The majority of branches contain individual members contributed by distinct PCR amplifications. The rare deletions and duplications are indicated by the prefix "x" and "d" respectively. Arrows highlight two mutations (a substitution at position 264 yielding a stop codon and a duplication at position 184) whose position within the tree implies that mutations can continue to accumulate following loss of functional heavy chain expression.

5 PCR artefacts make little contribution to the database of mutations; not only is the prevalence of nucleotide substitutions greatly in excess of that observed in control PCR amplifications ($<0.05 \times 10^{-3}$ bp $^{-1}$) but also identically mutated clones (as well as 10 dynamically related ones) are found in independent amplifications. In many cases, 15 generations within a lineage differ by a single nucleotide substitution indicating that only a small number of substitutions have been introduced in each round of mutation.

Analysis of V_{λ} rearrangements reveals that Ramos harbours an in-frame rearrangement 20 of $V_{\lambda}2.2-16$ (as described by Chapman et al. 1996) and an out-of-frame rearrangement of $V_{\lambda}2.2-25$. There is mutational diversity in both rearranged V_{λ} s although greater diversity has accumulated on the non-functional allele (Figure 1C).

A classic feature of antibody hypermutation is that mutations largely accumulate in the 25 V region but scarcely in the C. This is also evident in the mutations that have accumulated in the Ramos IgH locus (Figure 1D). M13 clones containing cDNA inserts extending through V_H , $C_{\mu}1$ and the first 87 nucleotides $C_{\mu}2$ are generated by PCR from the initial Ramos culture. The Pie charts (presented as in Figure 1A) depict the extent of mutation identified in the 341 nucleotide stretch of V_H as compared to a 380 nucleotide stretch of C_{μ} extending from the beginning of $C_{\mu}1$.

30 The IgM immunoglobulin produced by Ramos is present both on the surface of the cells and, in secreted form, in the culture medium. Analysis of the culture medium reveals

that Ramos secretes immunoglobulin molecules to a very high concentration, approximately 1 μ g/ml. Thus, Ramos is capable of secreting immunoglobulins to a level which renders it unnecessary to reclone immunoglobulin genes into expression cell lines or bacteria for production.

5

Example 2: V_H diversification in Ramos is constitutive

To address whether V gene diversification is ongoing, the cells are cloned and V_H diversity assessed using a MutS-based assay after periods of *in vitro* culture. The 10 Ramos V_H is PCR amplified and purified as described above using oligonucleotides containing a biotinylated base at the 5'-end. Following denaturation/renaturation (99°C for 3 min; 75 °C for 90 min), the extent of mutation is assessed by monitoring the binding of the mismatched heteroduplexed material to the bacterial mismatch-repair 15 protein MutS, filter-bound, with detection by ECL as previously described (Jolly et al., 1997).

The results indicate that V_H diversification is indeed ongoing (see Figure 2A). DNA is extracted from Ramos cells that have been cultured for 1 or 3 months following limit 20 dilution cloning. The rearranged V_H is PCR amplified using biotinylated oligonucleotides prior to undergoing denaturation/renaturation; mismatched heteroduplexes are then detected by binding to immobilised MutS as previously described (Jolly et al., 1997). An aliquot of the renatured DNA is bound directly onto membranes to confirm matched DNA loading (Total DNA control). Assays performed 25 on the Ramos V_H amplified from a bacterial plasmid template as well as from the initial Ramos culture are included for comparison.

The V_H genes are PCR amplified from Ramos cultures that have been expanded for four (Rc1) or six (Rc13 and 14) weeks (Figure 2B). A mutation rate for each clone is 30 indicated and is calculated by dividing the prevalence of independent V_H mutations at 4 or 6 weeks post-cloning by the presumed number of cell divisions based on a generation

time of 24 h. The sequences reveal step-wise mutation accumulation with a mutation rate of about 0.24×10^{-4} mutations bp⁻¹ generation⁻¹.

Direct comparison of the V_H mutation rate in Ramos to that in other cell-lines is not 5 straightforward since there is little information on mutation rates in other lines as judged by unselected mutations incorporated throughout the V_H obtained following clonal expansion from a single precursor cell. However, the prevalence of mutations following a two week expansion of 50 precursor BL2 cells has been determined under conditions of mutation induction (2.7×10^{-3} mutations bp⁻¹; Denépoux et al., 1997).
10 Similar experiments performed with Ramos under conditions of normal culture reveal a mutation prevalence of 2.3×10^{-3} mutations bp⁻¹. Various attempts to enhance the mutation rate by provision of cytokines, helper T cells etc. have proved unsuccessful. Thus, the rate of mutation that can be achieved by specific induction in BL2 cells appears to be similar to the constitutive rate of V_H mutation in Ramos.

15

Example 3: Examination of the nature of V_H mutations in Ramos

A database of mutational events is created which combines those detected in the initial 20 Ramos culture (from 141 distinct sequences) with those detected in four subclones that have been cultured in various experiments without specific selection (from a further 135 distinct sequences). This database is created after the individual sets of sequences have been assembled into dynastic relationships (as detailed in the legend to Figure 1B) to ensure that clonal expansion of an individual mutated cell does not lead to a specific mutational event being counted multiple times. Here an analysis of this composite 25 database of 340 distinct and presumably unselected mutational events (200 contributed by the initial Ramos culture and 140 from the expanded subclones) is described; separate analysis of the initial and subclone populations yields identical conclusions.

The overwhelming majority of the mutations (333 out of 340) are single nucleotide 30 substitutions. A small number of deletions (4) and duplications (3) are observed but no untemplated insertions; these events are further discussed below. There are only five

sequences which exhibited nucleotide substitutions in adjacent positions; however, in three of these five cases, the genealogy revealed that the adjacent substitutions have been sequentially incorporated. Thus, the simultaneous creation of nucleotide substitutions in adjacent positions is a rare event.

15 The distribution of the mutations along the V_H is highly non-random (See Figure 3). Independently occurring base substitutions are indicated at each nucleotide position. The locations of CDR1 and 2 are indicated. Nucleotide positions are numbered from the 3'-end of the sequencing primer with nucleotide position +1 corresponding to the 10 first base of codon 7; codons are numbered according to Kabat. Mutations indicated in italics (nucleotide position 15, 193, 195 and 237) are substitutions that occur in a mutated subclone and have reverted the sequence at that position to the indicated consensus.

15 The major hotspot is at the G and C nucleotides of the Ser82a codon, which has previously been identified as a major intrinsic mutational hotspot in other V_H genes (Wagner et al., 1995; Jolly et al., 1996) and conforms to the RGYW consensus (Rogozin and Kolchanov, 1992; Betz et al., 1993). Whilst the dominant intrinsic mutational hotspot in many V_H genes is at Ser31, this codon is not present in the Ramos 20 consensus V_H (or its germline counterpart) which have Gly at that position. The individual nucleotide substitutions show a marked bias in favour of transitions (51% rather than randomly-expected 33%). There is also a striking preference for targeting G and C which account for 82% of the nucleotides targeted (Table 1).

Table 1. Nucleotide substitution preferences of hypermutation in Ramos

		Frequency of substitution to				
		T	C	G	A	Total
5	Parental nucleotide					
T		3.9		1.2	3.0	8.1
10	C	17.4		12.6	4.8	34.8
G		7.2	15.9		24.0	47.1
15	A	2.4	1.8	5.7	-	9.9

Single nucleotide substitutions were computed on the V_H coding strand and are given as the percentage of the total number (333) of independent, unselected nucleotide substitutions identified.

Example 4: Selection of hypermutating cells by IgM-loss

Analysis of the Ramos variants reveals several mutations that must have inactivated V_H (see Figure 1B) suggesting it might be possible for the cells to lose IgM expression but remain viable. If this is the case, Ig expression loss would be an easy means to select a constitutively hypermutating B cell line.

Analysis of the Ramos culture reveals it to contain 8% surface IgM⁻ cells. Such IgM⁻ loss variants are generated during in vitro culture, as follows. The starting Ramos culture is transfected with a pSV2neo plasmid, diluted into 96-well plates and clones growing in selective medium allowed to expand. Flow cytometry performed on the

expanded clones six months after the original transfection reveals the presence of IgM-loss variants, constituting 16% and 18% of the two clonal populations (Rc13 and Rc14) shown here (Figure 4A). Enrichment by a single round of sorting yields subpopulations that contain 87% (Rc13) and 76% (Rc14) surface IgM-negative cells. Following PCR amplification of the rearranged V_H gene in these subpopulations, sequencing reveals that 75% (Rc13) and 67% (Rc14) of the cloned V_H segments contained a nonsense (stop), deletion (del) or duplication (dup) mutation within the 341 nucleotide V_H stretch analysed. The remainder of the clones are designated wild type (wt) although no attempt is made to discriminate possible V_H -inactivating missense mutations. The 4 deletions and 3 duplications identified in the Rc13 population are all distinct whereas only 4 distinct mutations account for the 7 Rc14 sequences determined that harbour deletions. The nature of the deletions and duplications is presented in Figure 6: each event is named with a letter followed by a number. The letter gives the provenance of the mutation (A, B and C being the cloned TdT⁻ control transfectants, D, E and F the TdT⁺ transfectants and U signifies events identified in the initial, unselected Ramos culture); the number indicates the first nucleotide position in the sequence string. Nucleotides deleted are specified above the line and nucleotides added (duplications or non-templated insertions) below the line; single nucleotide substitutions are encircled with the novel base being specified. The duplicated segments of V_H origin are underlined; non-templated insertions are in bold. With several deletions or duplications, the event is flanked by a single nucleotide of unknown provenance. Such flanking changes could well arise by nucleotide substitution (rather than non-templated insertion) and these events therefore separately grouped; the assignment of the single base substitution (encircled) to one or other end of the deletion/duplication is often arbitrary.

25

The IgM⁻ cells are enriched in a single round of sorting prior to PCR amplification and cloning of their V_H segments. The sequences reveal a considerable range of V_H -inactivating mutations (stop codons or frameshifts) (Figure 4) although diverse inactivating mutations are even evident in IgM-loss variants sorted after only 6 weeks of clonal expansion (see Figure 5). In Figure 5A expression of TdT in three pSV-p β G/TdT and three control transfectants of Ramos is compared by Western blot analysis

of nuclear protein extracts. Nalm6 (a TdT-positive human pre-B cell lymphoma) and HMy2 (a TdT-negative mature human B lymphoma) provided controls.

In Figure 5B, pie charts are shown depicting independent mutational events giving rise to IgM-loss variants. IgM⁻ variants (constituting 1-5% of the population) are obtained

5 by sorting the three TdT⁺ and three TdT⁻ control transfectants that have been cultured for 6 weeks following cloning. The V_H regions in the sorted subpopulations are PCR amplified and sequenced. The pie charts depict the types of mutation giving rise to V_H inactivation with the data obtained from the TdT⁺ and TdT⁻ IgM⁻ subpopulations separately pooled. Abbreviations are as in Figure 4A except that "ins" indicates clones 10 containing apparently non-templated nucleotide insertions. Clones containing deletions or duplications together with multiple nucleotide non-templated insertions are only included within the "ins" segment of the pie. Only unambiguously distinct mutational events are computed. Thus, of the 77 distinct V_H-inactivating mutations identified in the TdT⁺ IgM-loss subpopulations, 30 distinct stop codon mutations are identified; if 15 the same stop codon have been independently created within the IgM-loss population derived from a single Ramos transfectant, this would have been underscored.

The stop codons are created at variety of positions (Figure 4B) but are not randomly located. Figure 4B summarises the nature of the stop codons observed in the Rc13 and 20 Rc14 IgM-loss populations. At least eight independent mutational events yield the nonsense mutations which account for 20 out of the 27 non-functional V_H sequences in the Rc13 database; a minimum of ten independent mutational events yield the nonsense mutations which account for 15 of the 22 non-functional V_H sequences in the Rc14 database. The numbers in parentheses after each stop codon give the number of 25 sequences in that database that carry the relevant stop codon followed by the number of these sequences that are distinct, as discriminated on the basis of additional mutations. Analysis of stop codons in IgM-loss variants selected from four other clonal populations reveals stop codon creation at a further five locations within V_H. In data obtained in six independent experiments, stop codon creation is restricted to 16 of the 39 possible sites; 30 the DNA sequences at these preferred sites being biased (on either coding or non-coding strand) towards the RGYW consensus.

Not surprisingly, whereas deletions and insertions account for only a small proportion of the mutations in unselected Ramos cultures (see above), they make a much greater contribution when attention is focused on V_H -inactivating mutations. It is notable that a large proportion of the IgM-loss variants can be accounted for by stop-codon/frameshift mutations in the V_H itself. This further supports the proposal that hypermutation in Ramos is preferentially targeted to the immunoglobulin V domain - certainly rather than the C domain or, indeed other genes (such as the Ig α /Ig β sheath) whose mutation could lead to a surface IgM $^-$ phenotype. It also may well be that the Ramos V_H is more frequently targeted for hypermutation than its productively rearranged V_{λ} , a conclusion supported by the pattern of mutations in the initial culture (Figure 1C).

Selection of cells by detection of Ig loss variants is particularly useful where those variants are capable of reverting, i.e. of reacquiring their endogenous Ig-expressing ability. The dynasty established earlier (Figure 1B) suggests not only that IgM-loss cells could arise but also that they might undergo further mutation. To confirm this, IgM-loss variants sorted from Rc13 are cloned by limiting dilution. Three weeks after cloning, the presence of IgM $^+$ revertants in the IgM $^-$ subclones is screened by cytoplasmic immunofluorescence analysis of 5×10^4 cells; their prevalence is given (Figure 4C). These IgM $^+$ revertants are then enriched in a single round of sorting and the V_H sequences of the clonal IgM $^-$ variant compared to that of its IgM $^+$ revertant descendants.

Cytoplasmic immunofluorescence of ten expanded clonal populations reveals the presence of IgM $^+$ revertants at varying prevalence (from 0.005% to 1.2%; Figure 4C) allowing a mutation rate of 1×10^{-4} mutations bp $^{-1}$ generation $^{-1}$ to be calculated by fluctuation analysis. This is somewhat greater than the rate calculated by direct analysis of unselected mutations (0.25×10^{-4} mutations bp $^{-1}$ generation $^{-1}$; see above), probably in part reflecting that different IgM-loss clones revert at different rates depending upon the nature of the disrupting mutation. Indeed, the sequence surrounding the stop codons in the IgM-loss derivatives of Rc13 reveals that TAG32 conforms well to the RGYW

consensus (R = purine, Y = pyrimidine and W = A or T; Rogozin and Kolchanov, 1992) which accounts for a large proportion of intrinsic mutational hotspots (Betz et al., 1993) whereas TAA33 and TGA36 do not (Figure 4D).

5 **Example 5: Selection of a novel Ig binding activity**

In experiments designed to demonstrate development of novel binding affinities, it is noted that most members of the Ramos cell line described below express a membrane IgM molecule which binds anti-idiotype antibodies (anti-Id1 and anti-Id2), specifically

10 raised against the Ramos surface IgM. However, a few cells retain a surface IgM, yet fail to bind the anti-idiotype antibody. This is due to an alteration in binding affinity in the surface IgM molecule, such that it no longer binds antibody. Cells which express a surface IgM yet cannot bind antibody can be selected in a single round of cell sorting according to the invention.

15

This is demonstrated by isolating μ positive/id-negative clones which have lost the capacity to bind to anti-Id2 despite the retention of a surface IgM, by ELISA. The clones are sequenced and in six independent clones a conserved V_H residue, K70, is found to be mutated to N, M or R as follows:

20

Clone	Mutation
2	K70N AAG-AAC
	S77N AGC-AAC
4	K70M AAG-ATG
9	S59R AGT-AGG
	K70N AAG-AAC
10	K70N AAG-AAC
12	K70N AAG-AAC
13	K70R AAG-AGG

No mutations were observed in the light chain. Thus, it is apparent that mutants may be selected from the Ramos cell line in which the Ig molecule produced has a single base-pair variation with respect to the parent clone.

5 Making use of an anti-Id1, a similar population of cells is isolated which retain expression of the Ig μ constant region but which have lost binding to the anti-idiotype antibody. These cells are enriched by sorting cytometry and the sequence of V_H determined (Figure 7). This reveals six mutations when compared with the consensus sequence of the starting population. Two of these mutations result in amino acid 10 sequence changes around CDR3 (R->T at 95 and P->H at 98). Thus, selection of more subtle changes in the immunoglobulin molecule are selectable by assaying for loss of binding.

15 In further experiments, hypermutating cells according to the invention are washed, resuspended in PBS/BSA (10^8 cells in 0.25ml) and mixed with an equal volume of PBS/BSA containing 10% (v/v) antigen-coated magnetic beads. In the present experiment, streptavidin coated magnetic beads (Dynal) are used. After mixing at 4° C on a roller for 30 mins, the beads are washed three times with PBS/BSA, each time bringing down the beads with a magnet and removing unbound cells. remaining cells 20 are then seeded onto 96 well plates and expanded up to 10^8 cells before undergoing a further round of selection. Multiple rounds of cell expansion (accompanied by constitutively-ongoing hypermutation) and selection are performed. After multiple rounds of selection, the proportion of cells which bind to the beads, which is initially at or close to background levels of 0.02%, begins to rise.

25

After 4 rounds, enrichment of streptavidin binding cells is seen. This is repeated on the fifth round (Figure 8). The low percentage recovery reflects saturation of the beads with cells since changing the cell:bead ratio from vast excess to 1:2 allows a recovery of approximately 20% from round five streptavidin binding cells (Figure 9). This 30 demonstrates successful selection of a novel binding specificity from the hypermutating Ramos cell line, by four rounds of iterative selection.

Nucleotide sequencing of the heavy and light chains from the streptavidin binding cells predicts one amino acid change in V_H CDR3 and four changes in V_L (1 in FR1, 2 in CDR1 and 1 in CDR2) when compared with the consensus sequence of the starting 5 population (Figure 11).

To ensure that the binding of streptavidin is dependent on expression of surface immunoglobulin, immunoglobulin negative variants of the streptavidin binding cells are enriched by sorting cytometry. This markedly reduces the recovery of streptavidin 10 binding cells with an excess of beads. The cells recovered by the Dynal-streptavidin beads from the sorted negative cells are in fact Ig μ positive and most likely represent efficient recovery of Ig μ streptavidin binding cells contaminating the immunoglobulin negative sorted cell population.

15 Preliminary data suggest that the efficiency of recovery is reduced as the concentration of streptavidin on the beads is reduced (Figure 9). This is confirmed by assaying the recovery of streptavidin binding cells with beads incubated with a range of concentrations of streptavidin (Figure 10). The percentage of cells recoverable from a binding population is dictated by the ratio of beads to cells. In this experiment the ratio 20 is < 1: 1 beads:cells.

In a further series of experiments, a further two rounds of selection are completed, taking the total to 7. This is accomplished by reducing the concentration of streptavidin bound to the beads from 50 μ g/ml in round 5 to 10 μ g/ml in round 7. Although the 25 secretion levels of IgM is comparable for the populations selected in rounds 4 to 7 (Figure 12), streptavidin binding as assessed by ELISA is clearly greatly increased in rounds 6 and 7, in comparison with round 4 (Figure 13).

This is confirmed by assessment of binding by Surface Plasmon Resonance on a 30 BiaCore chip coated with streptavidin (Figure 14). The supernatant from round 7 is injected to flow across the chip at point A, and stopped at point B. At point C, anti-

human IgM is injected, to demonstrate that the material bound to the streptavidin is IgM. The gradient A-B represents the association constant, and the gradient B-C to dissociation constant. From the BiaCore trace it is evident that round 6 supernatant displays superior binding characteristics to that isolated from round 4 populations or

5. unselected Ramos cells.

Antibodies from round 6 of the selection process also show improved binding with respect to round 4. Binding of cells from round 6 selections to streptavidin-FITC aggregates, formed by preincubation of the fluorophore with a biotinylated protein, can 10 be visualised by FACS, as shown in Figure 15. Binding to round 4 populations, unselected Ramos cells or IgM negative Ramos is not seen, indicating maturation of streptavidin binding.

Use of unaggregated streptavidin-FITC does not produce similar results, with the 15 majority of round 6 cells not binding. This, in agreement with ELISA data, suggests that binding to streptavidin is due to avidity of the antibody binding to an array of antigen, rather than to a monovalent affinity. Higher affinity binders may be isolated by sorting for binding to non-aggregated streptavidin-FITC.

20 In order to determine the mutations responsible for the increased binding seen in round 6 cells over round 4 cells, the light and heavy chain antibody genes are amplified by PCR, and then sequenced. In comparison with round 4 cells, no changes in the heavy chain genes are seen, with the mutation R103S being conserved. In the light chain, mutations V23F and G24C are also conserved, but an additional mutation is present at 25 position 46. Wild-type Ramos has an Aspartate at this position, whilst round 6 cells have an Alanine. Changes at this position are predicted to affect antigen binding, since residues in this region contribute to CDR2 of the light chain (Figure 16). It seems likely that mutation D46A is responsible for the observed increase in binding to streptavidin seen in round 6 cells.

Example 6: Construction of transgene comprising hypermutation-directing sequences

5

It is known that certain elements of Ig gene loci are necessary for direction of hypermutation events *in vivo*. For example, the intron enhancer and matrix attachment region Ei/MAR has been demonstrated to play a critical role (Betz *et al.*, 1994). Moreover, the 3' enhancer E3' is known to be important (Goyenechea *et al.*, 1997).

10 However, we have shown that these elements, whilst necessary, are not sufficient to direct hypermutation in a transgene.

In contrast, provision of Ei/MAR and E3' together with additional J_K-C_K intron DNA and C_K is sufficient to confer hypermutability. A βG-C_K transgene is assembled by 15 joining an 0.96 Kb PCR-generated KpnI-SpeI β-globin fragment (that extends from -104 with respect to the β-globin transcription start site to +863 and has artificial KpnI and SpeI restriction sites at its ends) to a subfragment of L_KΔ[3'F1] [Betz *et al.*, 1994] that extends from nucleotide 2314 in the sequence of Max *et al* [1981] through Ei/MAR, C_K and E3', and includes the 3'F1 deletion.

20

Hypermutation is assessed by sequencing segments of the transgene that are PCR amplified using Pfu polymerase. The amplified region extends from immediately upstream of the transcription start site to 300 nucleotides downstream of J_K5.

25

This chimeric transgene is well targeted for mutation with nucleotide substitutions accumulating at a frequency similar to that found in a normal Ig_K transgene. This transgene is the smallest so far described that efficiently recruits hypermutation and the results indicate that multiple sequences located somewhere in the region including and flanking C_K combine to recruit hypermutation to the 5'-end of the β-globin/Ig_K chimaera.

30

The recruitment of hypermutation can therefore be solely directed by sequences lying towards the 3'-end of the hypermutation domain. However, the 5'-border of the mutation domain in normal Ig genes in the vicinity of the promoter, some 100-200 nucleotides downstream of the transcription start site. This positioning of the 5'-border of the mutation domain with respect to the start site remains even in the β G-C κ transgene when the β -globin gene provides both the promoter and the bulk of the mutation domain. These results are consistent with findings made with other transgenes indicating that it is the position of the promoter itself that defines the 5'-border of the mutation domain.

10

The simplest explanation for the way in which some if not all the κ regulatory elements contribute towards mutation recruitment is to propose that they work by bringing a hypermutation priming factor onto the transcription initiation complex. By analogy with the classic studies on enhancers as transcription regulatory elements, the Ig κ enhancers may work as regulators of hypermutation in a position and orientation-independent manner. Indeed, the data obtained with the β G-C κ transgene together with previous results in which E3' was moved closer to C κ [Betz *et al.*, 1994] reveal that the hypermutation-enhancing activity of E3' is neither especially sensitive to its position or orientation with respect to the mutation domain.

20

Ei/MAR normally lies towards the 3'-end of the mutation domain. Whilst deletion of Ei/MAR drastically reduces the efficacy of mutational targeting, its restoration to a position upstream of the promoter (and therefore outside the transcribed region) gives a partial rescue of mutation but without apparently affecting the position of the 5'-border of the mutational domain. Independent confirmation of these results was obtained in transgenic mice using a second transgene, *tk-neo::C κ* , in which a *neo* transcription unit (under control of the HSV*tk* promoter) is integrated into the C κ exon by gene targeting in embryonic stem cells [Zou, *et al.*, 1995]. In this mouse, following V κ -J κ joining, the Ig κ Ei/MAR is flanked on either side by transcription domains: the V gene upstream and *tk::neo* downstream. The *tk-neo* gene is PCR amplified from sorted germinal centre B cells of mice homozygous for the *neo* insertion.

For the *tk-neo* insert in *tk-neo*::C_K mice, the amplified region extends from residues 607 to 1417 [as numbered in plasmid pMCNeo (GenBank accession U43611)], and the nucleotide sequence determined from position 629 to 1329. The mutation frequency of 5 endogenous VJ_K rearrangements in *tk-neo* ::C_K mice is determined using a strategy similar to that described in Meyer *et al.*, 1996. Endogenous VJ_{K5} rearrangements are amplified using a V_K FR3 consensus forward primer (GGACTGCAGTCAGGTTCA GTGGCAGTGGG) and an oligonucleotide L_KFOR [Gonzalez-Fernandez and Milstein, (1993) PNAS (USA) 90:9862-9866] that primes 10 back from downstream of the J_K cluster.

Although the level of mutation of the *tk-neo* is low and it is certainly less efficiently targeted for mutation than the 3'-flanking region of rearranged V_K genes in the same 15 cell population, it appears that - as with normal V genes - the mutation domain in the *neo* gene insert starts somewhat over 100 nucleotides downstream of the transcription start site despite the fact that Ei/MAR is upstream of the promoter.

Thus, transgenes capable of directing hypermutation in a constitutively hypermutating cell line may be constructed using Ei/MAR, E3' and regulatory elements as defined 20 herein found downstream of J_K. Moreover, transgenes may be constructed by replacement of or insertion into endogenous V genes, as in the case of the *tk-neo* ::C_K mice, or by linkage of a desired coding sequence to the J_K intron, as in the case of the β G- 25 C_K transgene.

References

Adetugbo, K., Milstein, C. and Secher, D.S. (1977). Molecular analysis of spontaneous somatic mutants. *Nature* 265, 299-304.

Bahler, D.W. and Levy, R. (1992). Clonal evolution of a follicular lymphoma: Evidence for antigen selection. *Proc. Natl. Acad. Sci. USA* 89, 6770-6774.

Bass, S., R. Greene, and J.A. Wells. (1990). Hormone Phage: An Enrichment Method for Variant Proteins With Altered Binding Properties. *Proteins*. 8, 309-314.

Braeuninger, A., Küppers, R., Strickler, J.G., Wacker, H.-H., Rajewsky, K. and Hansmann, M.-L. (1997). Hodgkin and Reed-Sternberg cells in lymphocyte predominant disease represent clonal populations of germinal center-derived tumor B cells. *Proc. Natl. Acad. Sci. USA* 94, 9337-9342.

Berek, C. and Milstein, C. (1988). The dynamic nature of the antibody repertoire. *Immunol. Rev.* 105, 5-26.

Betz, A.G., Neuberger, M.S. and Milstein, C. (1993). Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. *Immunol. Today* 14, 405-411.

Betz, A. G., Milstein, C., Gonzalez-Fernandez, A., Pannell, R., Larson, T. and Neuberger, M. S., *Cell* 1994. 77: 239-248.

Bonfield, J.K., Smith, K.F. and Staden, R. (1995). A new DNA-sequence assembly program. *Nucleic Acids Res.* 23, 4992-99.

Brüggemann, M., Radbruch, A. and Rajewsky, K. (1982). Immunoglobulin V region variants in hybridoma cells. I. Isolation of a variant with altered idiotypic and antigen binding specificity. *EMBO J.* 1, 629-634.

Capizzi and Jameson, (1973) *Mutat. Res.* 17:147-8

Chapman, C.J., Mockridge, C.I., Rowe, M., Rickinson, A.B. and Stevenson, F.K. (1995). Analysis of VH genes used by neoplastic B cells in endemic Burkitt's lymphoma shows somatic hypermutation and intraclonal heterogeneity. *Blood* 85, 2176-2181.

Chapman, C.J., Zhou, J.X., Gregory, C., Rickinson, A.B. and Stevenson F.K. (1996). VH and VL gene analysis in sporadic Burkitt's lymphoma shows somatic hypermutation, intraclonal heterogeneity and a role for antigen selection. *Blood* 88, 3562-3568.

Chapman, K.B. and Szostak, J.W. (1994) *Curr. op. Struct. Biol.*, 4, 618-622.

Chui, Y.-L., Lozano, F., Jarvis, J.M., Pannell, R. and Milstein, C. (1995). A reporter gene to analyse the hypermutation of immunoglobulin genes. *J. Mol. Biol.* 249, 555-563.

Clackson, T. and Wells, J.A. (1994) *Trends Biotechnol.* 12, 173-84.

Coffino, P. and Scharff, M.D. (1971). Rate of somatic mutation in immunoglobulin production by mouse myeloma cells. *Proc. Natl. Acad. Sci. USA* 68, 219-223.

Cull, M.G., Miller, J.F. and Schatz, P.J. (1992) *Proc Natl Acad Sci U S A* 89, 1865-9.

Denépoux, S., Razanajaona, D., Blanchard, D., Meffre, G., Capra, J.D., Banchereau, J. and Lebecque, S. (1997). Induction of somatic mutation in a human B cell line *in vitro*. *Immunity* 6, 35-46.

Drake, J.W. (1998) *Genetics* 148:1667-1686.

Ellington, A.D. and Szostak, J.W. (1990) *Nature*, 346, 81822.

Gold, L., Poliskiy, B., Uhlenbeck, O. and Yarus, M. (1995) *Annu Rev Biochem.* 64, 763-97.

Goossens, T., Klein, U. and Küppers, R. (1998). Frequent occurrence of deletions and duplications during somatic hypermutation: Implications for oncogenic translocations and heavy chain disease. *Proc. Natl. Acad. Sci. USA* 95, 2463-2468.

Goyenechea, B., Klix, N., Yélamos, J., Williams, G. T., Riddell, A., Neuberger, M.S. and Milstein, C., *EMBO J.* 1997. 16, in the press.

Green, N.S., Lin, M.M. and Scharff, M.D. (1998). Immunoglobulin hypermutation in cultured cells. *Immunol. Rev.* 162, 77-87.

Green, R. and Szostak, J.W. (1992) *Science*, 258, 1910-5.

Jain, R., Roncella, S., Hashimoto, S., Carbone, A., Franco di Celle, P., Foa, R., Ferrarini, M and Chiorazzi, N. (1994). A potential role for antigen selection in the clonal evolution of Burkitt's lymphoma. *J. Immunol.* 153, 45-52.

Jolly, C.J., Klix, N. and Neuberger, M.S. (1997). Rapid methods for the analysis of immunoglobulin gene hypermutation: application to transgenic and gene-targeted mice. *Nucleic Acids Res.* 25, 1913-1919.

Jolly, C.J., Wagner, S.D., Rada, C.A., Klix, N., Milstein, C. and Neuberger, M.S. (1996). The targeting of somatic hypermutation. *Semin. Immunol.* 8, 159-168.

Joyce, G.F. (1994) *Curr. op. Structural Biol.*, 4, 331-336.

Klein, G., Giovannella, B., Westman, A., Stehlin, J. and Mumford, D. (1975). An EBV negative cell line established from a American Burkitt lymphoma; receptor characteristics, EBV infectability and permanent conversion into EBV-positive

sublines by in vitro infection. *Intervirology* 5, 319-334.

10 Klix *et al.*, (1998) *Eur. J. Immunol.* 28:317-326.

Lundberg, K.S., *et al.*, (1991) *Gene* 108:1-6.

Luria and Delbreck, (1943) *Genetics* 28:491-511.

15 McCafferty, J., Griffiths, A.D., Winter, G. and Chiswell, D.J. (1990) *Nature*, 348, 552-4.

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W. and Weigert, M. (1984). Generation of antibody diversity in the immune response of BALB/c mice to influenza virus haemagglutinin. *Proc. Natl. Acad. Sci. USA* 81, 3180-3184.

10 Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 9022-6.

Matthias, *et al.*, (1989) *NAR* 17, 6418.

Max, E. E., Maizel, J. V. and Leder, P., *J. Biol. Chem.* 1981, 256: 5116-5120.

Moore, M.J. (1995) *Nature*, 374, 766-7.

15 Neuberger, M.S. and Milstein, C. (1995). Somatic hypermutation. *Curr. Opin. Immunol.* 7, 248-254.

Parham, P. (ed). (1998). Somatic hypermutation of immunoglobulin genes. *Immunological Reviews*, Vol. 162 (Copenhagen, Denmark: Munksgaard).

Ratech, H. (1992). Rapid cloning of rearranged immunoglobulin heavy chain genes from 20 human B-cell lines using anchored polymerase chain reaction. *Biochem. Biophys. Res. Commun.* 182, 1260-1263.

Rogozin, I.B. and Kolchanov, N.A. (1992). Somatic hypermutation of immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochem. Biophys. Acta* 1171, 11-18.

25 Sharpe *et al.*, (1991) *EMBO J.* 10:2139-2145.

Smith, G.P. (1985) *Science*, 228, 1315-7.

Tomlinson, I.M. (1997). V Base database of human antibody genes. Medical Research Council, Centre for Protein Engineering, UK. <http://www.mrc-cpe.cam.ac.uk/>

Tuerk, C. and Gold, L. (1990) *Science*, 249, 505-10.

30 Wabl, M., Burrows, P.D., von Gabain, A. and Steinberg, C.M. (1985). Hypermutation at the immunoglobulin heavy chain locus in a pre-B-cell line. *Proc. Natl. Acad. Sci. USA* 82, 479-482.

Wagner, S.D., Milstein, C. and Neuberger, M.S. (1995). Codon bias targets mutation. *Nature* 376, 732.

Weill, J.-C. and Reynaud, C.-A. (1996). Rearrangement/hypermutation/gene conversion: when, where and why? *Immunol. Today* 17, 92-97.

5 Wilson, P.C., de Bouteiller, O., Liu, Y-J., Potter, K., Banchereau, J., Capra, J.D. and Pascual, V. (1998). Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J. Exp. Med.* 187, 59-70.

Wu, H., Pelkonen, E., Knuutila, S and Kaartinen M. (1995). A human follicular lymphoma B cell line hypermutates its functional immunoglobulin genes in vitro. *Eur. J. Immunol.* 25, 3263-3269.

10 Wu, H., and Kaartinen, M., (1995) *Scand. J. Immunol.* 42:52-59.

Zhu, M., Rabinowitz, J.L., Green, N.S., Kobrin, B.J. and Scharff, M.D. (1995) A well-differentiated B cell line is permissive for somatic mutation of a transfected immunoglobulin heavy-chain gene. *Proc. Natl. Acad. Sci. USA* 92, 2810-2814.

15 Zou, X., Xian, J., Popov, A. V., Rosewell, I. R., Muller, M. and Brüggemann, M., *Eur. J. Immunol.* 1995. 25: 2154-62.

Claims

1. A method for preparing a lymphoid cell line capable of directed constitutive hypermutation of a target nucleic acid region, comprising screening a cell population for ongoing target sequence diversification, and selecting a cell in which the rate of target nucleic acid mutation exceeds that of other nucleic acid mutation by a factor of 100 or more.
2. A method according to claim 1, wherein the lymphoid cell line is derived from an immunoglobulin-expressing cell.
3. A method according to claim 1 or claim 2, wherein the lymphoid cell line is derived from or related to a cell type which hypermutates *in vivo*.
4. A method according to claim 3, wherein the cell line is a Burkitt lymphoma, follicular lymphoma or diffuse large cell lymphoma cell line.
5. A method according to any preceding claim, further comprising the steps of isolating one or more cells which display target sequence diversification, and comparing the rate of accumulation of mutations in the target sequences with that in non-target sequences in the isolated cells.
6. A method according to any preceding claim, wherein the target sequence is an immunoglobulin V-gene sequence.
7. A method according to claim 6, wherein the cells are screened by assessing loss of an expressed immunoglobulin.
8. A method according to any one of claims 1 to 6, wherein the cells are screened by assessment of mutation rates by direct sequencing of the target sequences.

9. A method according to any one of claims 1 to 6, wherein the cells are screened by an immunofluorescence technique.

10. A method according to any preceding claim, wherein the rate of mutation in the cell is modulated by the administration of a mutagen or the expression of a sequence-modifying gene product.

11. A method for preparing a gene product having a desired activity, comprising the steps of:

- 10 a) expressing a nucleic acid encoding the gene product in a population of cells according to claim 1, operably linked to a sequence which directs hypermutation;
- b) identifying a cell or cells within the population of cells which expresses a mutated gene product having the desired activity; and
- c) establishing one or more clonal populations of cells from the cell or cells identified in step (b), and selecting from said clonal populations a cell or cells which expresses a gene product having an improved desired activity.

12. A method according to claim 11, wherein the cell or cells direct constitutive hypermutation to an endogenous V gene locus.

20 14. A method according to claim 12 or claim 13, wherein the control sequences which direct hypermutation are selected from sequences occurring downstream of a J gene cluster.

25 15. A method according to claim 14, wherein the control sequences comprise elements Ei/MAR, C_K plus flanking regions and E3' as defined according to Klix *et al.*, (1998) Eur J. Immunol. 28:317-326.

30 16. A method according to any one of claims 11 to 15, wherein the nucleic acid region operatively linked to control sequences which direct hypermutation is an exogenous sequence inserted into the cell or cells.

17. A method according to claim 16, wherein the exogenous sequence comprises a heterologous coding sequence operably linked to control sequences homologous to the cell or cells which direct hypermutation.

5

18. A method according to claim 17, wherein an endogenous V region coding sequence is replaced by a heterologous coding sequence.

19. A method according to any one of claims 11 to 18, wherein the gene product is 10 an immunoglobulin.

20. A method according to any one of claims 11 to 19, wherein the gene product is a DNA binding protein.

15 21. A method according to any one of claims 11 to 20, wherein the desired activity is a binding activity.

22. A method according to any one of claims 11 to 18, wherein the gene product is an enzyme.

20 23. A method according to any one of claims 11 to 22, wherein steps b) and c) are iteratively repeated.

25. The use of a cell capable of directed constitutive hypermutation of a specific 25 nucleic acid region in the preparation of a gene product having a desired activity.

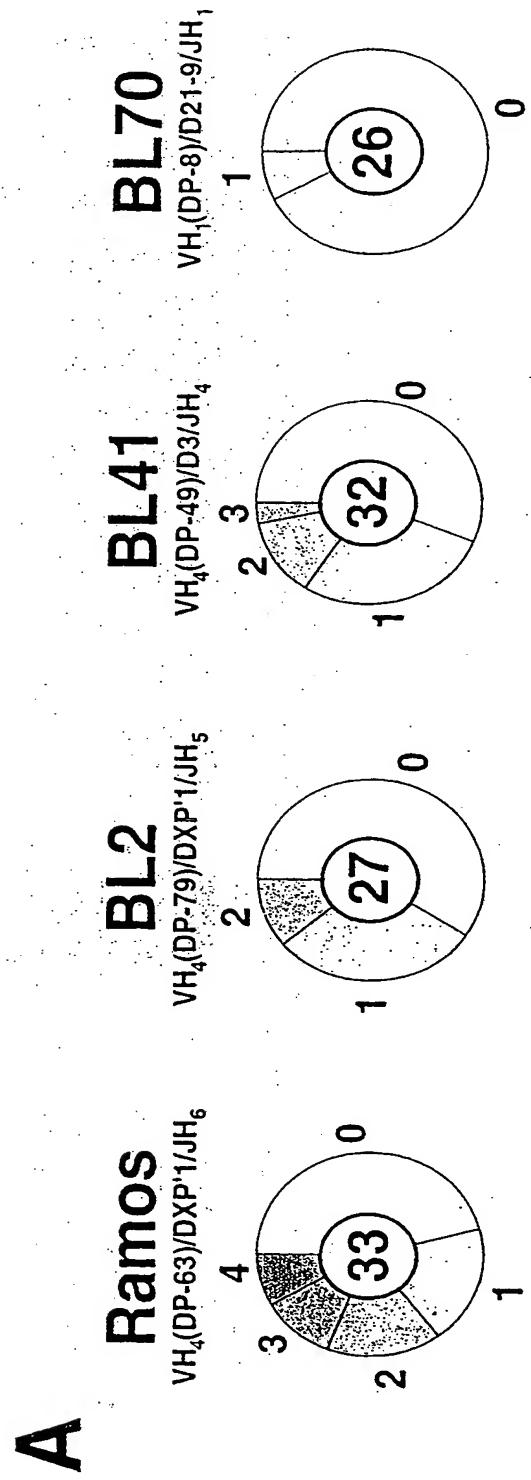


FIG. 1

11

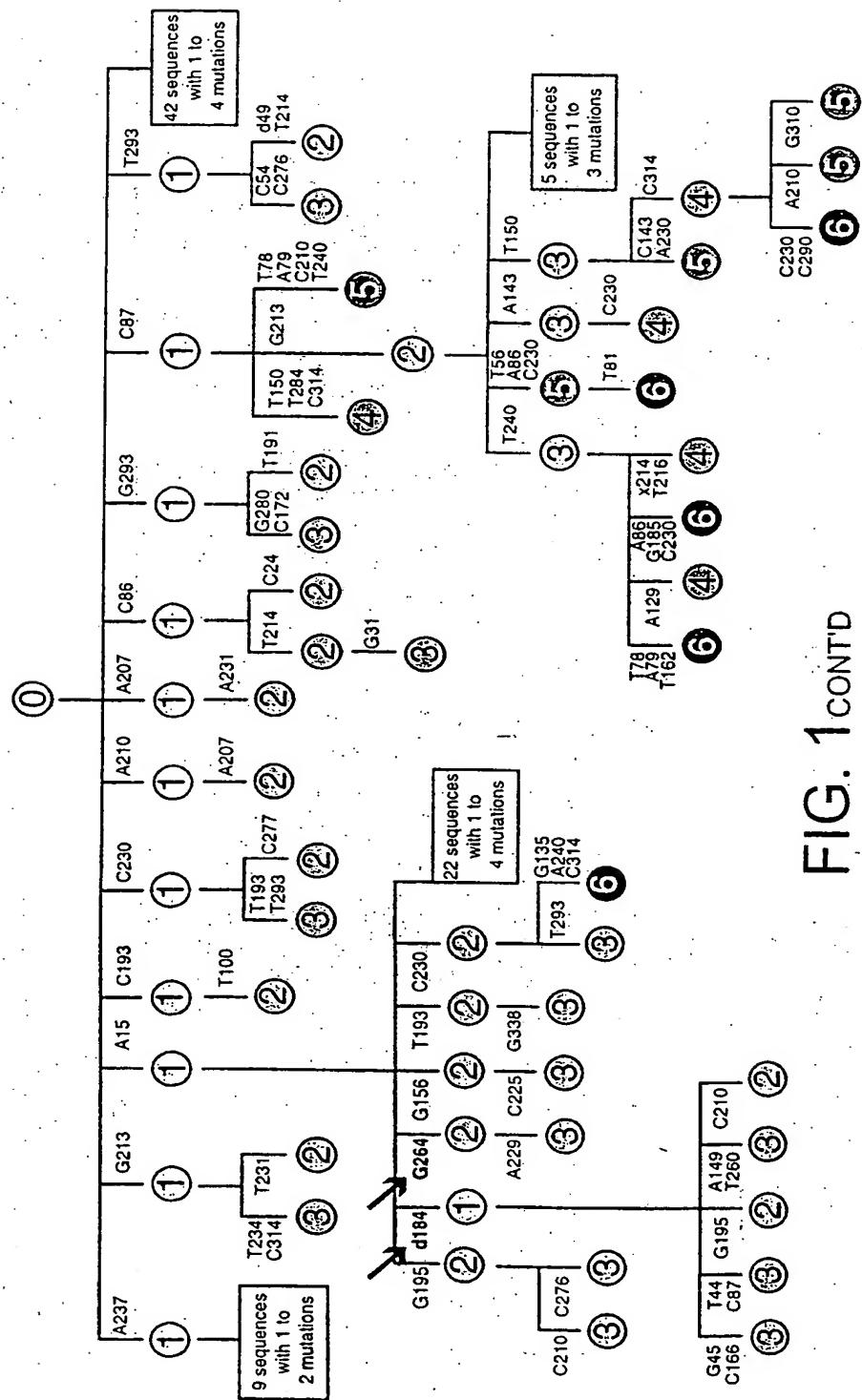


FIG. 1 (CONT'D)

3 / 20

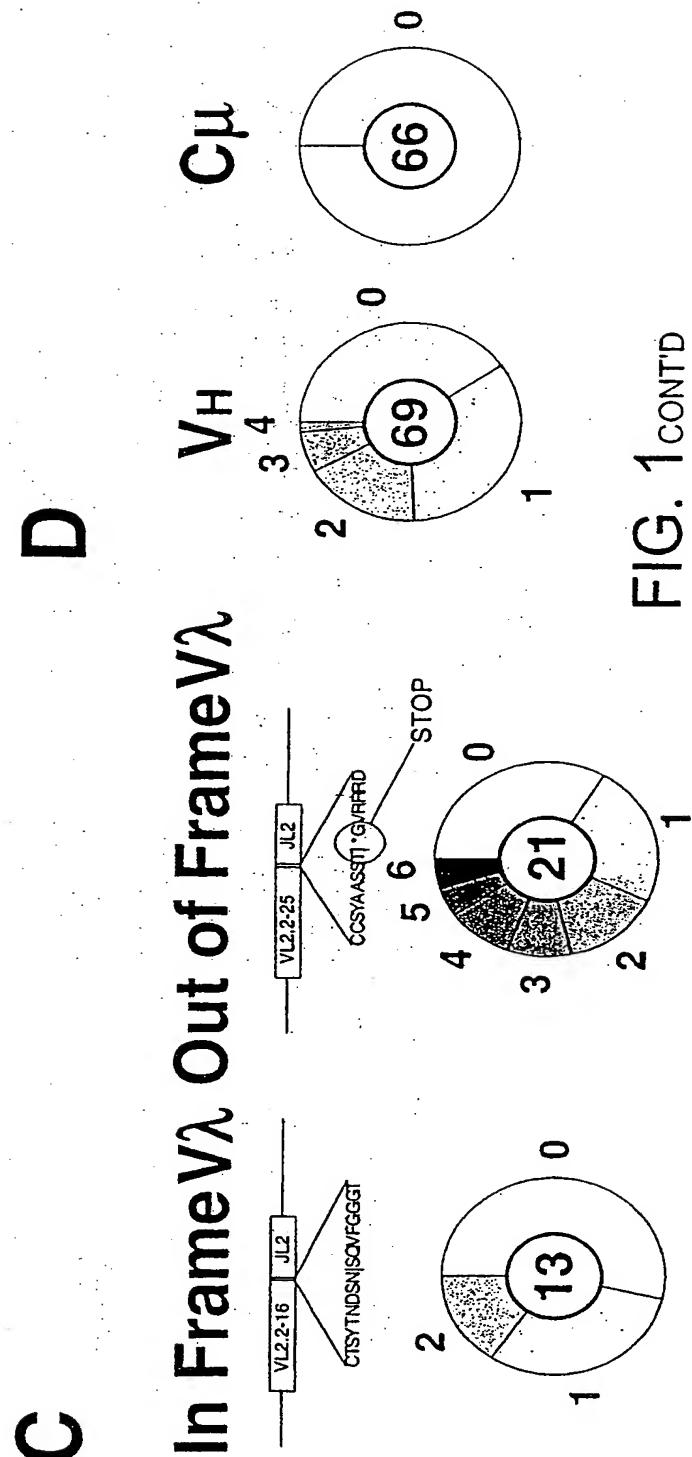
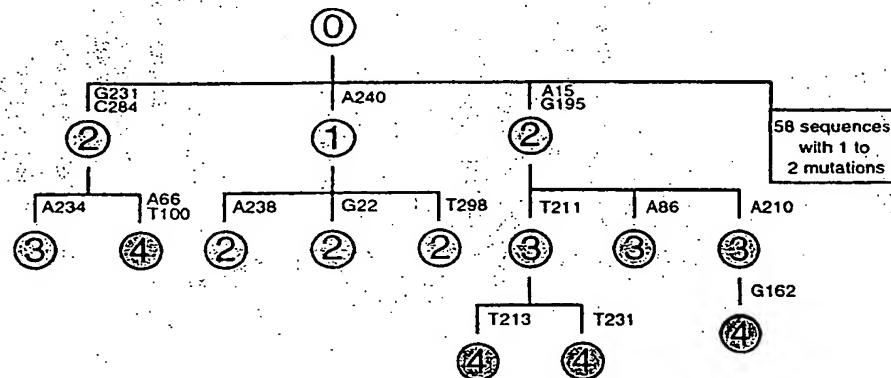
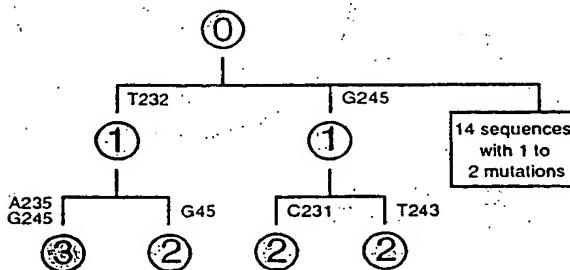
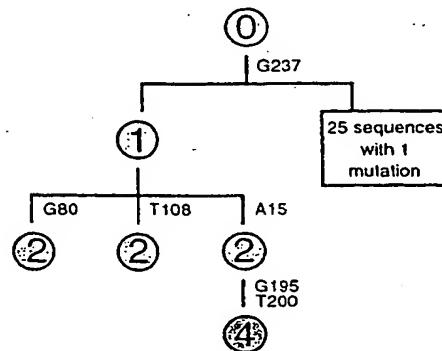


FIG. 1 CONT'D

BEST AVAILABLE COPY

A

	MutS bound DNA	Total DNA	mutations. bp10 ⁻³
V _H Plasmid			
Clone Rc1	1 Month		0.3
	3 Months		1.8
Initial Culture			3.0

B**Clone Rc13** 0.24×10^{-4} mutn.bp⁻¹.div⁻¹**Clone Rc14** 0.22×10^{-4} mutn.bp⁻¹.div⁻¹**Clone Rc1** 0.27×10^{-4} mutn.bp⁻¹.div⁻¹**FIG. 2**

SUBSTITUTE SHEET (RULE 26)

5 / 20

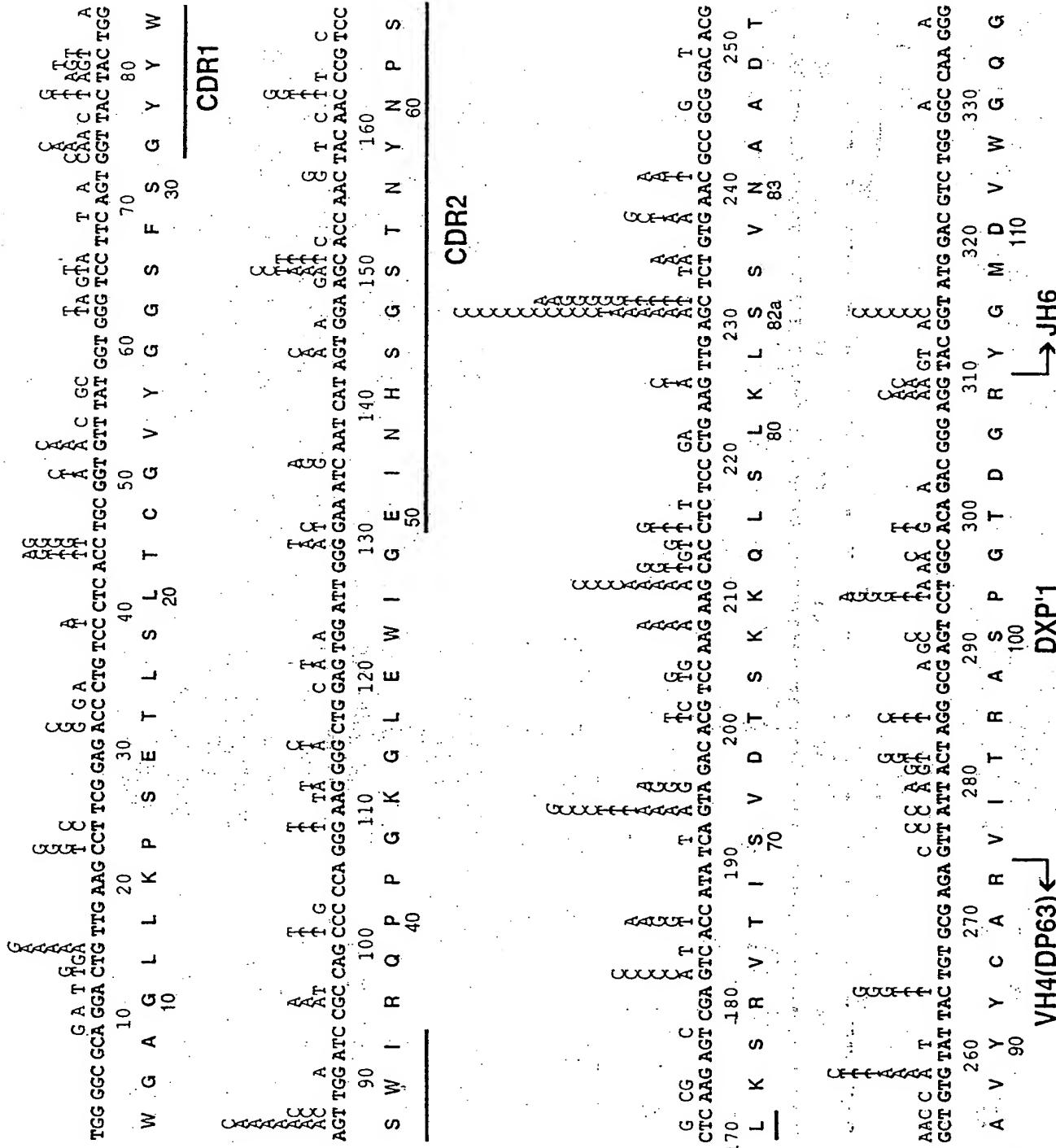


FIG. 3

6 / 20

BEST AVAILABLE COPY

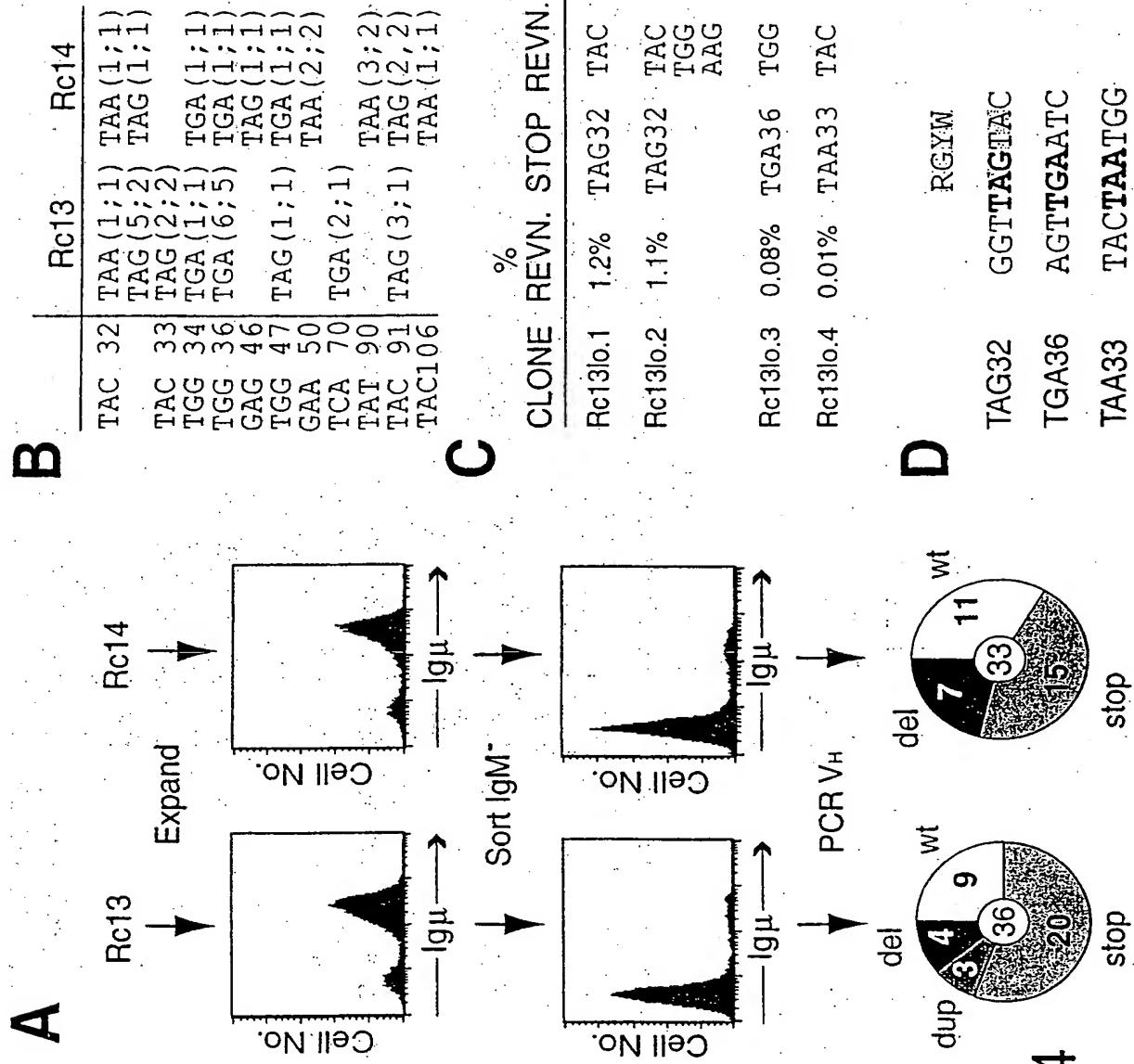


FIG. 4

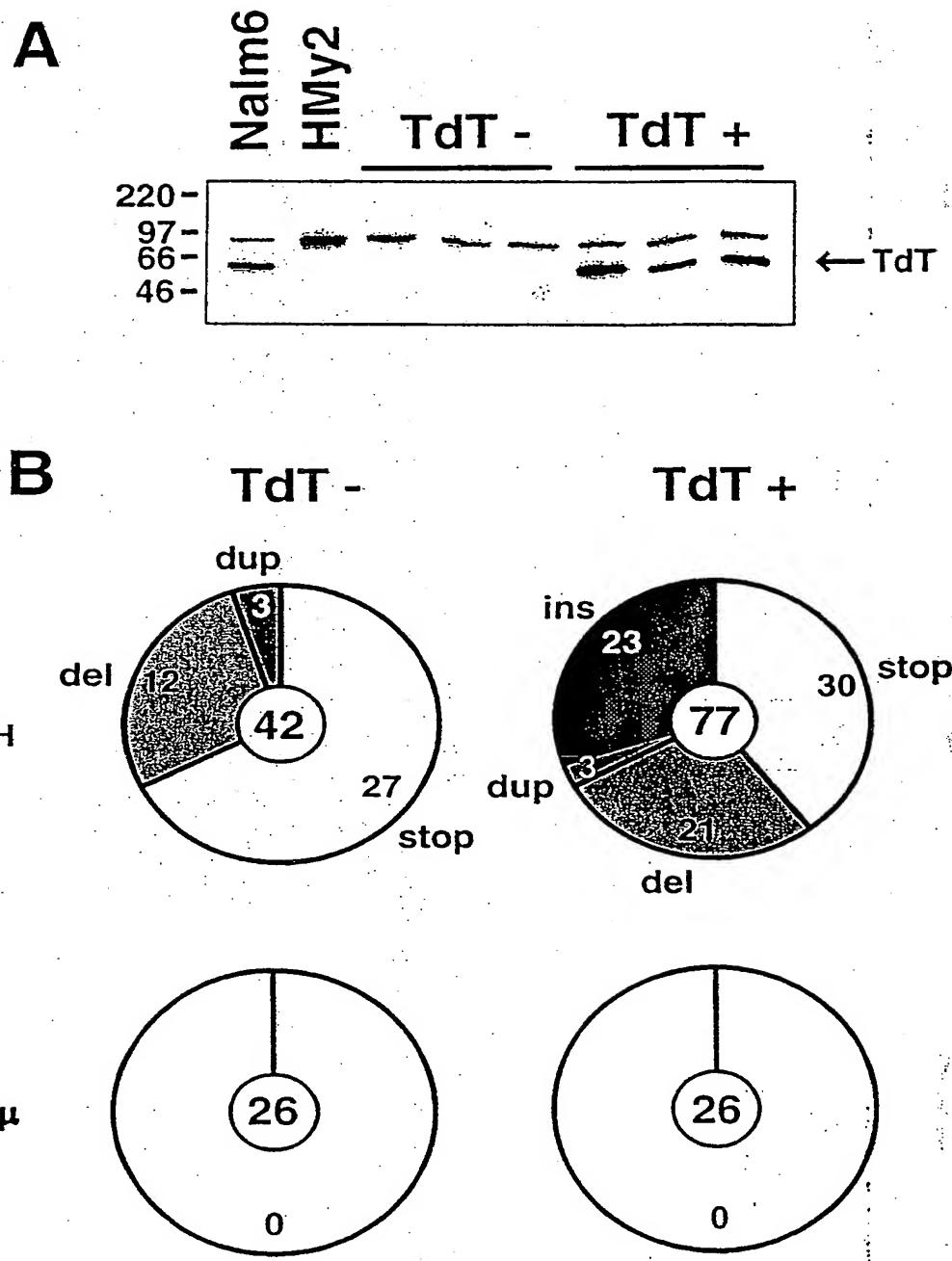


FIG. 5

TdT negative

TdT positive		TdT negative	
Insertion (+/- Del/Dup)		Deletion	
		D27 TCGAGACCTCTCTCTCGG	D3 CGGCC AGGACTCTTGAGCC
A62	GGTCCTTCAGCTGGTTACTA	D31 ACCCTA T CTCGG	D56 ATGCGGGG S Subp. CGGGG CGTCATG S Subp. CGGGG
A120	CTGGAT T GGGAA	D219 CCTCAA T TGAGC	D71 GTCGTC A CTACTG
A276	TATTA T TGTC, 18bp. TACT A GGCGC	D150 CACCA T TACAC	D75 TTCTAG C TGGAGT
A306	CGGGTA C GTGATG	D109 AAGGGG C TGGGT	D126 TGCGGA A TCATCAT T AGTGA
B93	CCGCC G CCCCCA	E38 CCCCT C CTGGGTGT	D223 AGTTG A GGCGC CTCTGTG
B98	AGCCCC C AGGAA	E81 CTGGAGTGGGA, 17bp. TCGAG T GGATT	D232 TCTCTGAGCCG G GGCCCGGCTCTGAGACGGCCG
B227	TGACCTCTG T AGGCC	E88 TGGATC G CCC AGCCCC	D235 GTCAG C GGAG G GGCGG
C82	TGAGT T TGA, 37bp. GAGT G TATTG	E93 CGGCCA G CCCCA	D252 GCTGTGATTA T TCTG G GGAGA
C209	AGACCC T CTCCCTAAGT T GAGTC	E136 ATCAT A TGCGAACCCACCA T AAACCC	D268 GCGAGA G GT A GGATATGAT
C187	ATATCA T AGAACGCCAGAA A ACCC	F66 CTTAG T GGTTACTACT G GGATT	D275 TTATTGG C TAGGG
U26	CGGAGA CC CTGTC	F183 T GGAT A ACACGT	D332 AACGA A GCAC
U199	AGTC CC AG A GGAC	F215 TCTCCCTGA, 18bp. CGGC C CGGAC	E3 GGGCG A GA A 18bp. CTTC AGTGT
U208	AGGGAG C GTGTC	F257 TGGCAG A TATTA	E51 TTTTT T GT T 15bp. TACT A TGTGAG
U268	GGAGA G TTATA T CTAGGG		E80 ACTG C CC C TTGAT
			E263 ACTGTC G GGAGTAA T ACT AGGGCG
			F69 GATCC G CCAGCC C AGGG A GGGG
			F168 CCTCA A AGATG G GT C ACCAT G
			F195 AGAC C CTCCAGAG A ACCTC
			F199 AGTCC A AGAG A ACCTGA
			F242 CGCGG A ACACGTG T TATCTGT G CGAGA
			F260 ATTACT T GG A CGAG
			F264 CTGTC G AGAG, 16bp. CGTC T GGGC
Events with flanking single nucleotide substitutions			
Deletion		Deletion	
		D45 ...CTCCCG T TATGGGGCTTCA	
B123	CTTGG T AAATC	D164 CGTCGCGAG A TGCA	
C109	AGGGG T GGAT	D216 CTCCCT A AG, 22bp. CGGA C ACCGC	
		E11 GACT T AGGCC	
		E54 TTATGGGG, 25bp. GTTG T TCCG	
		F188 TATCTGAGACAGCTGAGA G ACCT	
		F220 CTGGAGGAGCTCTGTG A GGCC	

FIG. 6

9 / 20

11/1 31/11
TGG GGC GCA GGA CTG TTG AAG CCT TCG GAGG ACC CTG TCC CTC ACC TGC GGT TAT GGT
W G A G L L K P S E T L S T C G V Y G

61/21 91/31
 GGG TCC TTC AGT GGT TAC TAC TGG AGC TGG ATC CGC CAG CCC CCA GGG AAG GGG CTG GAG
 G S F S G Y Y W W S W I R Q P P G K G L E
 AGT

1121/41 151/51
 TGG ATT GGG GAA ATC AAT CAT AGT GGA AGC ACC AAC TAC AAC CCG TCC CTC AAG AGT CGA
 W I G E I N H S G S T N Y N P S L K S R

181/61	211/71
GTC ACC ATA TCA GTA GAC ACG TCC AAG AAG CAG CTC TCC CTG AAG TTG AGC TCT GTG AAC	
V T I S V D T S K K H L S L K L S S V N	
ATC	CAC
M	H

241/81	271/91	
GCC GCG GAC ACG GCT GTG TAT TAC TGT GCG AGA GTT ATT ACT AGG GCG AGT CCT GGA ACA	A D T A V Y C A R V I T R A S P G T	
TCG S		CAT GGC H G

3301/101 331/111
GAC GGG AGG TAC GTC ATG GAC GTC TGG GGC CAA GGG ACC ACG
D G R Y G M D V W G Q G T T
GTT

FIG. 7

10 / 20

BEST AVAILABLE COPY

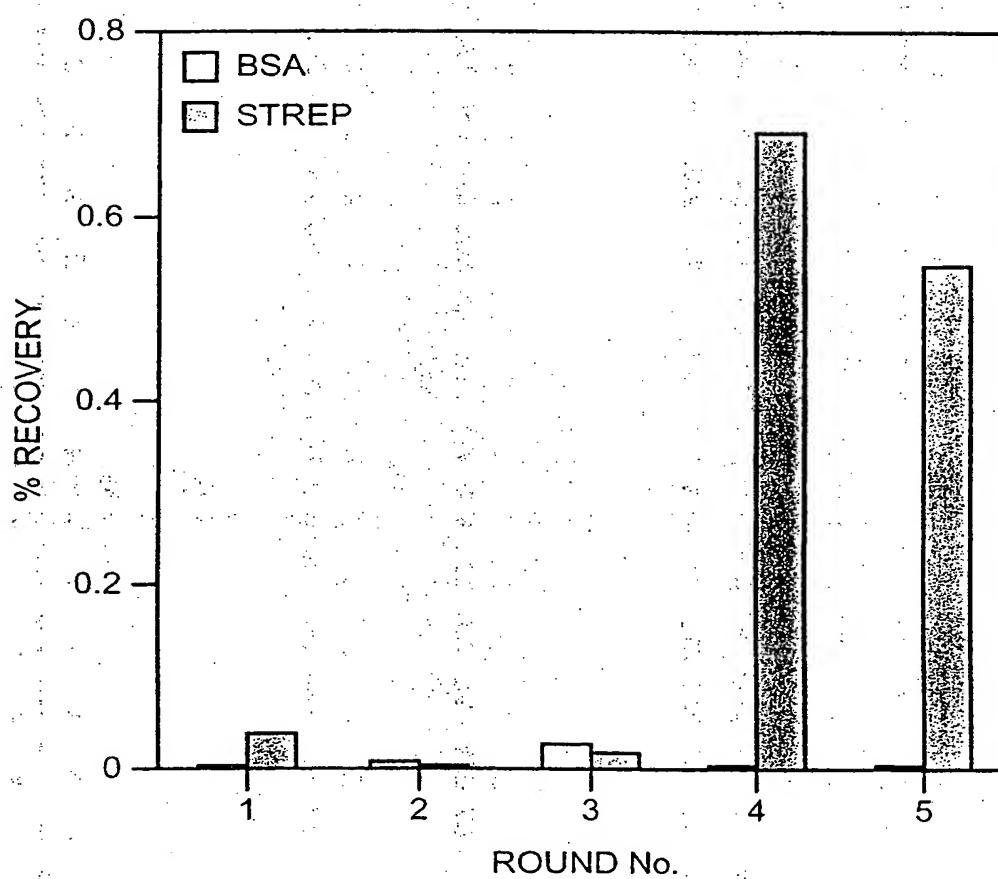
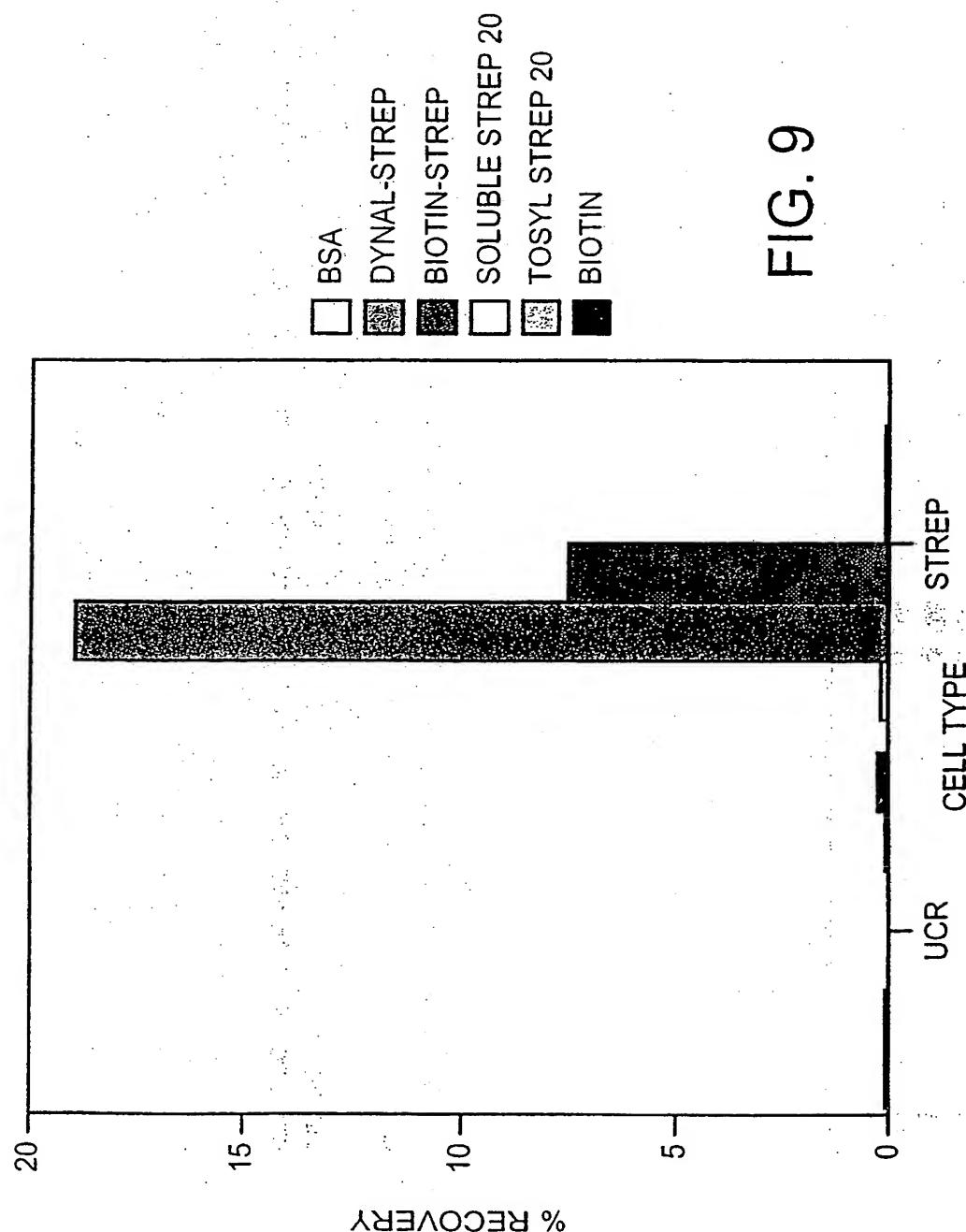


FIG. 8

11 / 20



12 / 20

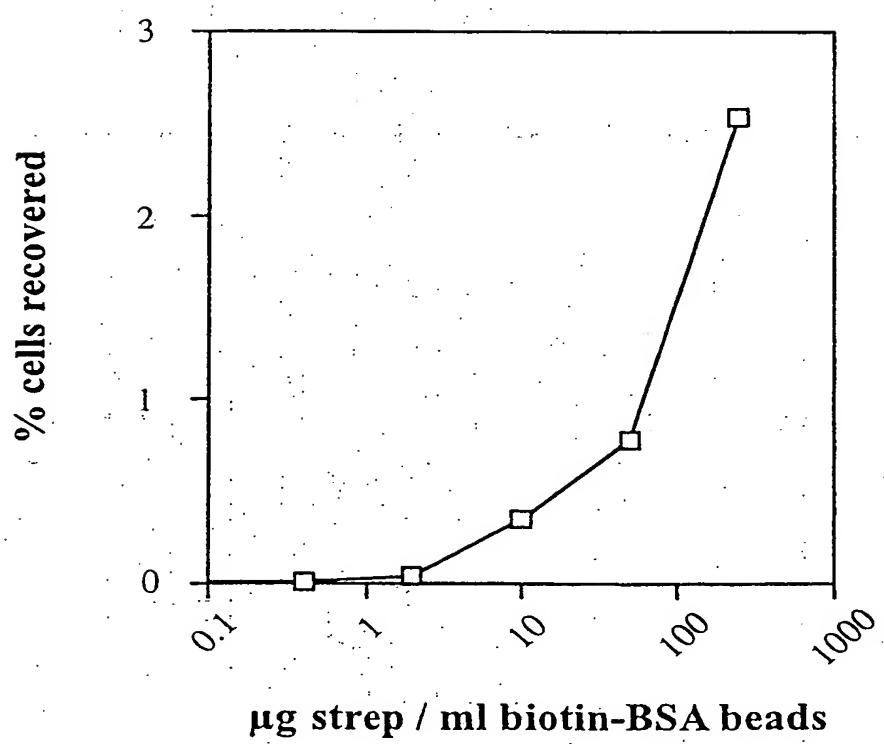


FIG. 10

FIG. 11

VII

1 / 1

31 / 11

31/11

TGG GGC GCA GGA CCT TTG AAG CCT TCG GAG ACC CTC TCC CTC ACC TGC GGT GTT TAT GGT
W G A G L K P S E T L S L T C G V Y G

61/21

21731
GGG TCC TTC AGT GGT TAC TAC TGG AGC TGG ATC CGC CAG CCC CCA GGG AAG GGG CTG GAG
G S F S G Y Y W S W I R Q P P G K G L E
GGA AGT AGT S

I

151 / 51

121/41 151/51
TGG ATT GGG GAA ATC AAT CAT AGT GGA ACC ACC AAC TAC AAC CCG TCC CTC AAG AGT CGA
W I G E I N H S G S T N Y N P S L K S R

181/61 211/71
 GTC ACC ATA TCA GTA GAC ACG TCC AAG CAG CTC TCC CTG AAG TTG AGC TCT GTG AAC
 V T I S V D T S K K H L S L K L S S V N
 CAC H AAC AAC N

241/81 271/91

GCC GAC ACG GCT GTG TAT TAC TGT GCG AGA GTT ATT ACT AGG GGT CCT GGA ACA
A A D T A V Y C A R V I T R A S P G T

301/101 331/111

VL

1/1

CCT GCC TCC GTG TCT GGG TCT CCT GGA CAG TCG ATC ACC ATC TCC TGC ACT GGA ACC AGC
 P A S V S G S P G Q S I T I S C T G T S
 TAT TAT Y

61/21

91/31

AGT GAC GTT GGT GGT TAT AAC TAT GTC TCC TGG TAC CAA AAC CCA GGC AAA GCC CCC
 S D V G G Y N Y V S W Y Q Q N P G K A P
 TTT TGT F C

121/41

151/51

AAA CTC ATG ATT TAT GAT GTC AGT AAT CGG CCC TCA GGG ATT TCT AAT CGC TTC TCT GGC
 K L M I Y D V S N R P S G I S N R F G S
 AAT N CGA TTA R L

181/61

211/71

TCC AAG TCT GGC AAC ACG GCC TCC CTG ACC ATC TCT GGG CTC CAG GCT GAC GAG GCT
 S K S G N T A S L T I S G L Q A D D E A
 ATC T

241/81

271/91

GAT TAT TAC TGC ACC TCA TAT ACA AAC GAC AGC AAT TCT CAG GTA TTC GGC GGA GGG ACC
 D Y C T S Y T N D S N S Q V F G G G T
 ACT T

FIG. 11 CONT'D

15 / 20

FIG. 12
IgM in supernatant

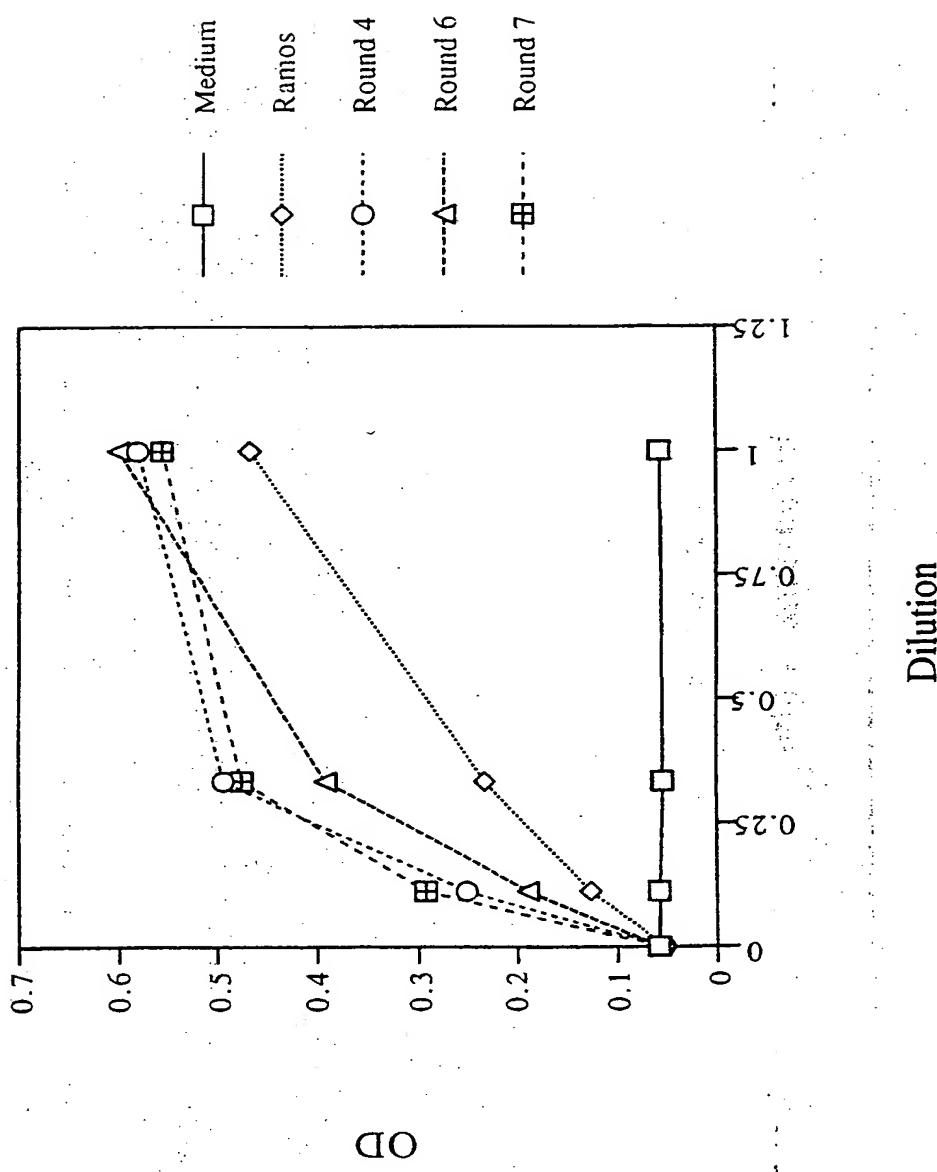
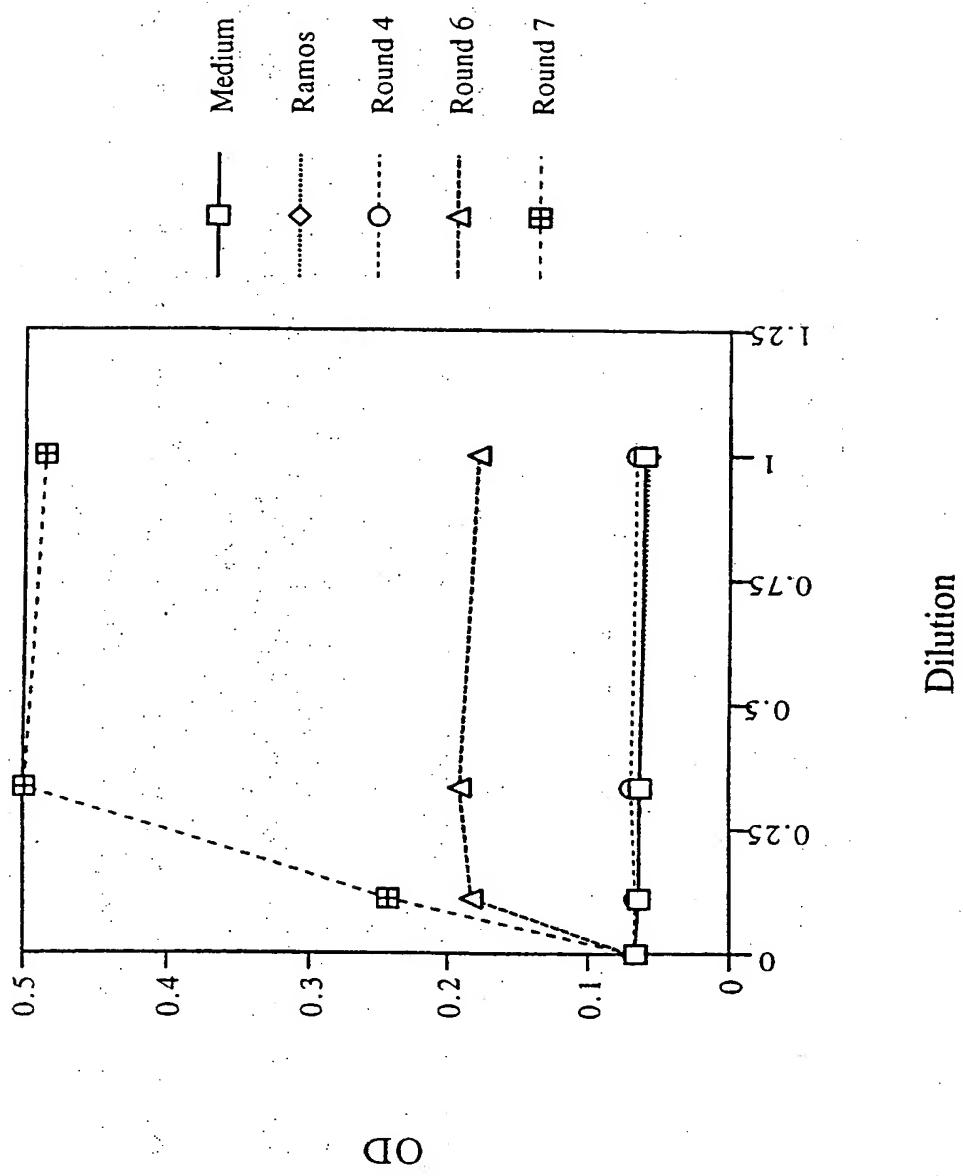


FIG. 13

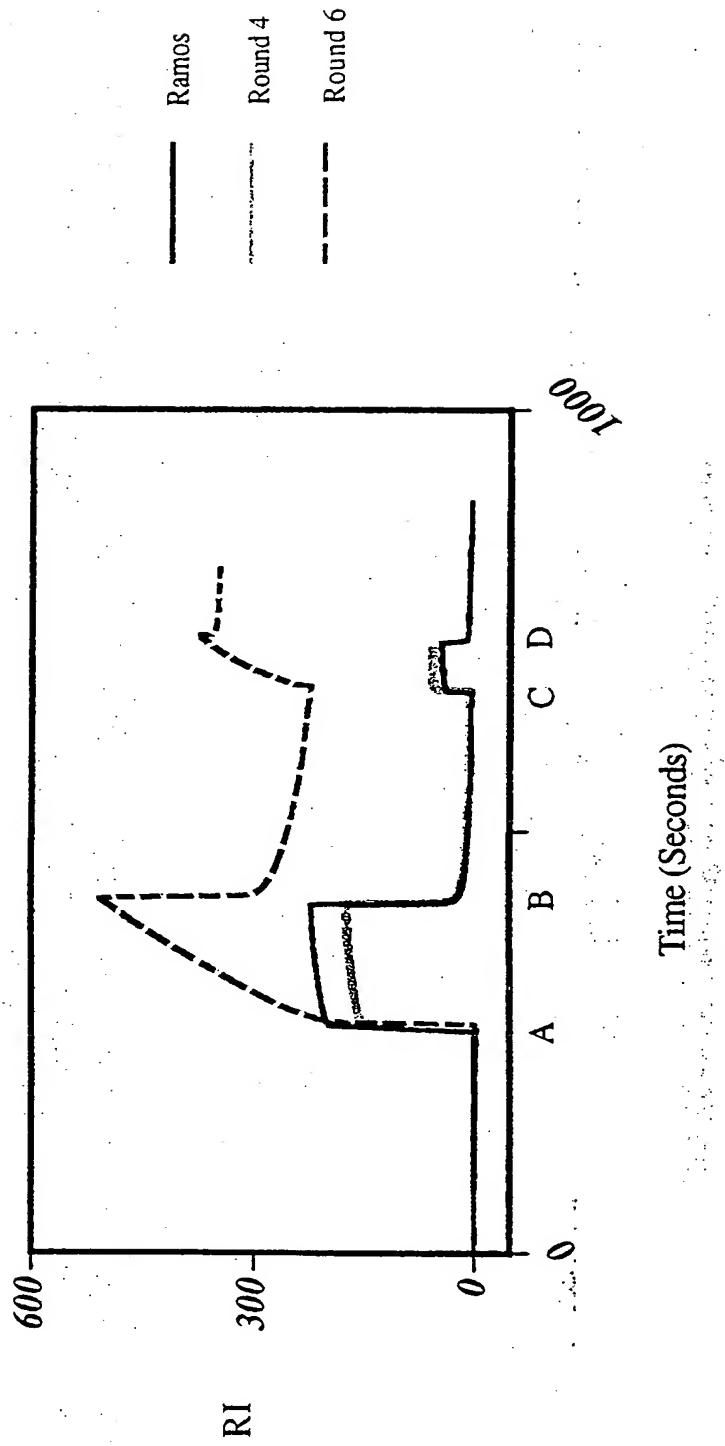
Streptavidin binding of Supernatants: ELISA



17 / 20

Streptavidin binding of Supernatants: Biacore

FIG. 14

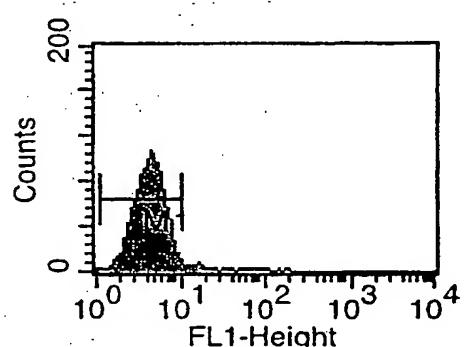
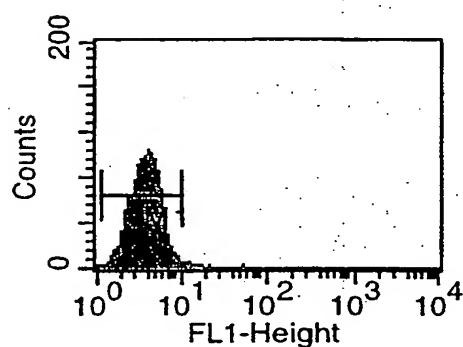


18 / 20

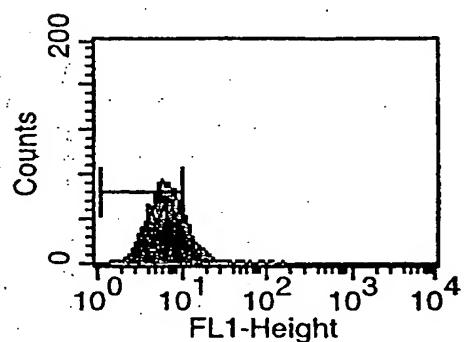
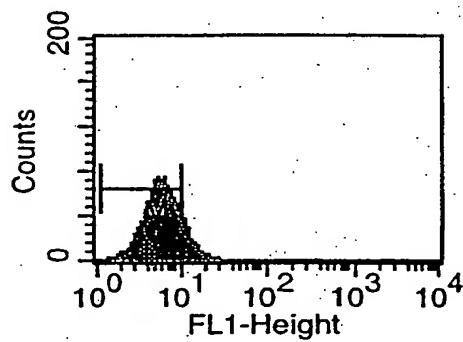
CD8-bio 1/500

CD8-bio 1/500
Strep-FITC 1/50

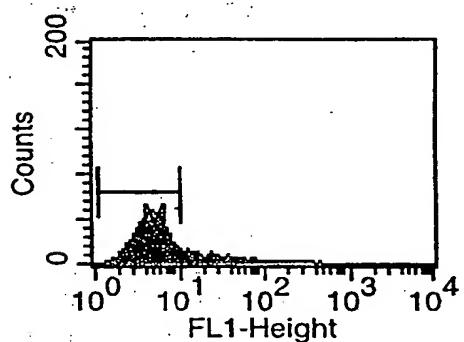
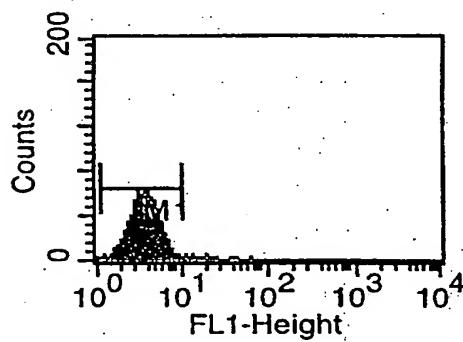
Ramos



Ramos
IgM -ve



Round 4



Round 6

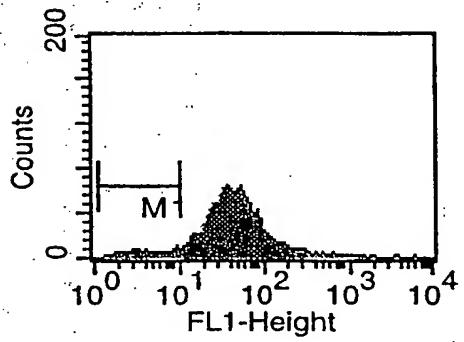
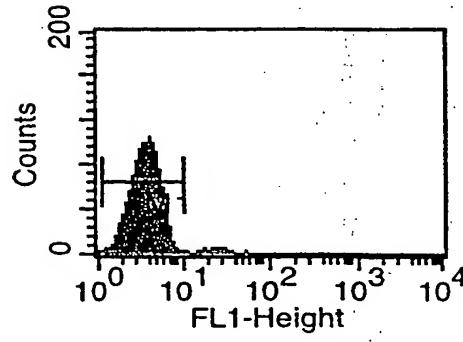


FIG. 15
SUBSTITUTE SHEET (RULE 26)

19 / 20

FIG. 16

VH

1/1	TGG GGC GCA GGA CTG TTG AAG CCT CCT TCG GAG ACC CTC TCC CTC ACC TGC GGT GTT TAT GGT											
W G A G L L K P S E T L S L T C G V Y G												
61/21	GGG TCC TTC AGT GGT TAC TAC TGG AGC TGG ATC CGC CAG CCC CCA GGG AAG GGG CTG GAG											
G S F S G Y Y W S W I R Q P P G K G L E												
CDR1	91/31											
121/41	TGG ATT GGG GAA ATC AAT CAT AGT GGA AGC ACC AAC TAC AAC CCG TCC CTC AAG AGT CGA											
W I G E I N H S G S T N Y N P S L K S R												
CDR2	151/51											
181/61	GTC ACC ATA TCA GTA GAC ACG TCC AAG AAG CAG CTC TCC TCC AAG TTG AGC TCT GTG AAC											
V T I S V D T S K K H L S L K S V N												
211/71	271/91 DJ											
241/81	GCC GCG GAC ACG GCT GTG TAT TAC TGT GCG AGA GTT ATT ACT AGG GCG AGT CCT GGA ACA											
A A D T A V Y Y C A R V I T R A S P G T												
301/101	331/111											
GAC GGG AGG TAC GGT ATG GAC GTC TGG GGC CAA GGG ACC ACG												
D G R Y G M D V W G Q G T T												
AGC												
S												

20 / 20

VL

1/1

CCT GCC TCC GTG TCT GGG TCT CCT GGA CAG TCG ATC ACC ATC TGC ACT GGA ACC AGC
 p A S V S G S P G Q S I T I S C T G T S

31/11

CDR1

61/21

AGT GAC GTT GGT TAT AAC TAT GTC TCC TGG TAC CAA AAC CCA GGC AAA GCC CCC
 S D V G G Y N Y V S W Y Q Q N P G K A P
 TTT TGT

F C

121/41

AAA CTC ATG ATT TAT GAT GTC AGT AAT CGG CCC TCA GGG ATT TCT AAT CGC TTC TCT GGC
 K L M I Y D V S N R P S G I S N R F G S
 GCT

A

CDR2

151/51

91/31

AGT GAC GTT GGT TAT AAC TAT GTC TCC TGG TAC CAA AAC CCA GGC AAA GCC CCC

S D V G G Y N Y V S W Y Q Q N P G K A P
 TTT TGT

181/61

TCC AAG TCT GGC AAC ACG GCC TCC CTG ACC ATC TCT GGG CTC CAG GCT GAC GAG GCT
 S K S G N T A S L T I S G L Q A D D E A

241/81

GAT TAT TAC TGC ACC TCA TAT ACA AAC GAC AGC AAT TCT CAG GTA TTC GGC GGA GGG ACC
 D Y Y C T S Y T N D S N S Q V F G G T

CDR3

271/91

FIG. 16 CONT'D

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 99/03358

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/10 C12N5/08
A01K67/027

C12N5/10

C07K16/00

C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q A61K A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	J. E. SALE AND M.S. NEUBERGER: "TdT-Accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line" IMMUNITY, vol. 9, no. 6, December 1998 (1998-12), pages 859-869, XP000876807 CELL PRESS, CAMBRIDGE, MA, US the whole document	1-25
Y	WO 93 12228 A (PASTEUR INSTITUT) 24 June 1993 (1993-06-24) the whole document	11-25 -/-

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the International filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the International filing date but later than the priority date claimed

"T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the International search

17 February 2000

Date of mailing of the International search report

07/03/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 851 epo nl.
Fax (+31-70) 340-3016

Authorized officer

Hornig, H

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/03358

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	C. J. JOLLY ET AL.: "Rapid methods for the analysis of immunoglobulin gene hypermutation: application to transgenic and gene targeted mice" NUCLEIC ACIDS RESEARCH, vol. 25, no. 10, 1997, pages 1913-1919, XP002130744 IRL PRESS LIMITED, OXFORD, ENGLAND cited in the application the whole document	11-25
Y	N. Klix ET AL.: "Multiple sequences from downstream of the J-kappa cluster can combine to recruit somatic hypermutation to a heterologous, upstream mutation domain" EUROPEAN J. IMMUNOLOGY, vol. 28, January 1998 (1998-01), pages 317-326, XP000876803 VCH VERLAGSGESELLSCHAFT MBH, WEINHEIM, BRD cited in the application the whole document	11-25
A	S. DENÉPOUX ET AL.: "Induction of somatic mutation in a human B cell line in vitro" IMMUNITY, vol. 6, January 1997 (1997-01), pages 35-46, XP000876832 CELL PRESS, CAMBRIDGE, MA, US cited in the application the whole document	
A	M. J. SHARPE ET AL.: "Somatic hypermutation of immunoglobulin kappa may depend on sequences 3' of c-kappa and occurs on passenger transgenes" EMBO J., vol. 10, no. 8, August 1991 (1991-08), pages 2139-2145, XP002130745 OXFORD UNIVERSITY PRESS, GB; cited in the application the whole document	
A	E. KÄLLBERG ET AL.: "Somatic mutation of immunoglobulin V genes in vitro" SCIENCE, vol. 271, 1 March 1996 (1996-03-01), pages 1285-1289, XP002130746 AAAS, WASHINGTON, DC, US the whole document	
		-/-

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 99/03358

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>M. MU-QUAN LIN ET AL.: "Sequence dependent hypermutation of the immunoglobulin heavy chain in cultured B cells" PROC. NATL. ACAD. SCI., vol. 94, May 1997 (1997-05), pages 5284-5289, XP002130747 NATL. ACAD. SCI., WASHINGTON, DC, US; the whole document</p>	
A	<p>WO 95 12689 A (GENE CHECK INC ;WAGNER ROBERT E JR (US)) 11 May 1995 (1995-05-11) the whole document</p>	
A	<p>US 5 750 335 A (GIFFORD DAVID K) 12 May 1998 (1998-05-12) the whole document</p>	

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int'l. Application No.

PCT/GB 99/03358

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
WO 9312228	A 24-06-1993	FR 2685004	A	18-06-1993	
		CA 2125192	A	24-06-1993	
		EP 0618965	A	12-10-1994	
		JP 7504086	T	11-05-1995	
		US 5756323	A	26-05-1998	
WO 9512689	A 11-05-1995	AU 700959	B	14-01-1999	
		AU 1050895	A	23-05-1995	
		AU 9607698	A	04-02-1999	
		CA 2175222	A	11-05-1995	
		EP 0733126	A	25-09-1996	
		JP 9504699	T	13-05-1997	
US 5750335	A 12-05-1998	WO 9322457	A	11-11-1993	

